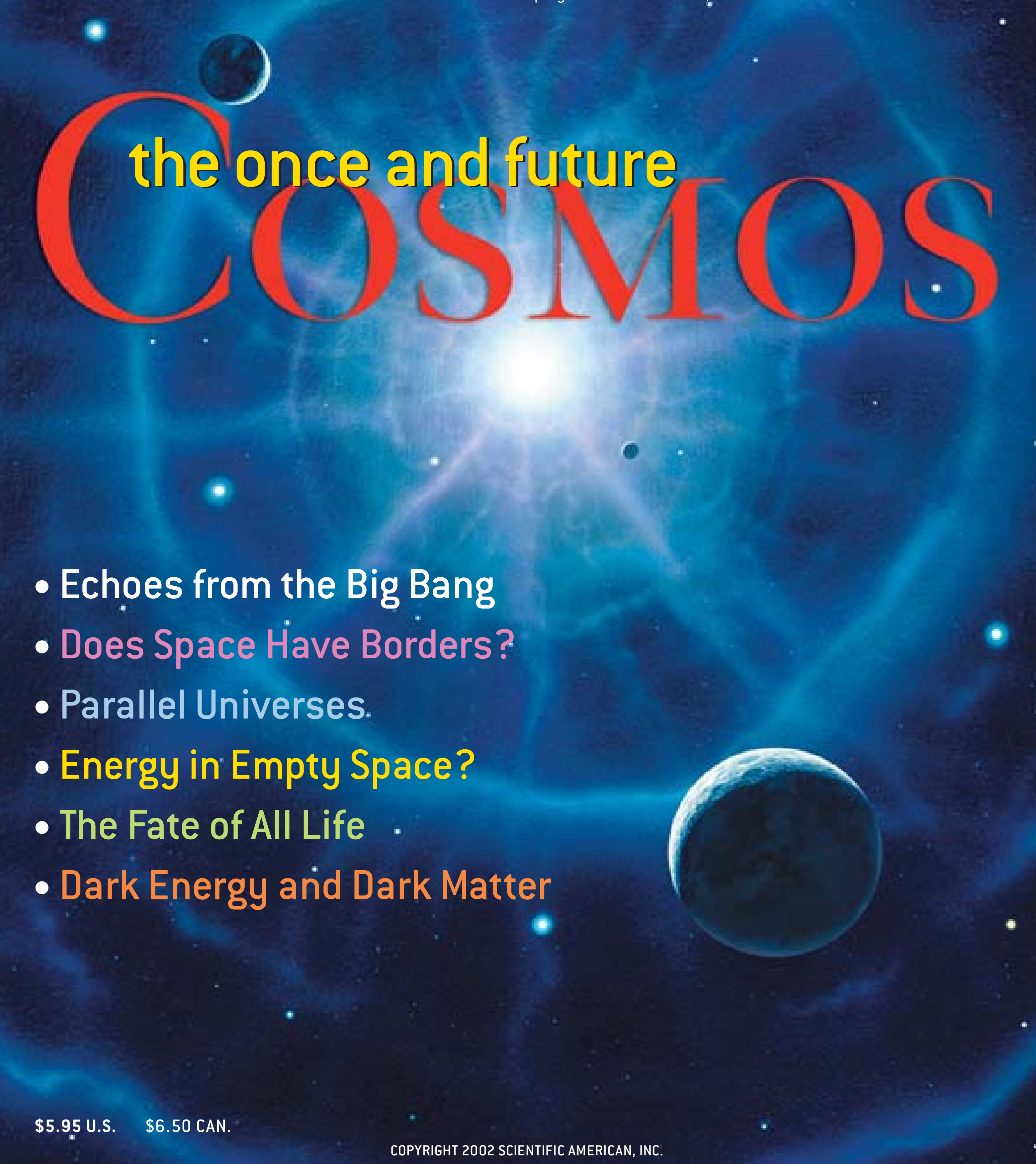
SPECIAL EDITION

SCIENTIFIC AMERICAN

WWW.SCIAM.COM

Display until December 31, 2002



the once and future COSMOS

- Echoes from the Big Bang
- Does Space Have Borders?
- Parallel Universes
- Energy in Empty Space?
- The Fate of All Life
- Dark Energy and Dark Matter

$5.95 U.S. $6.50 CAN.

contents 2002

the once and future cosmos

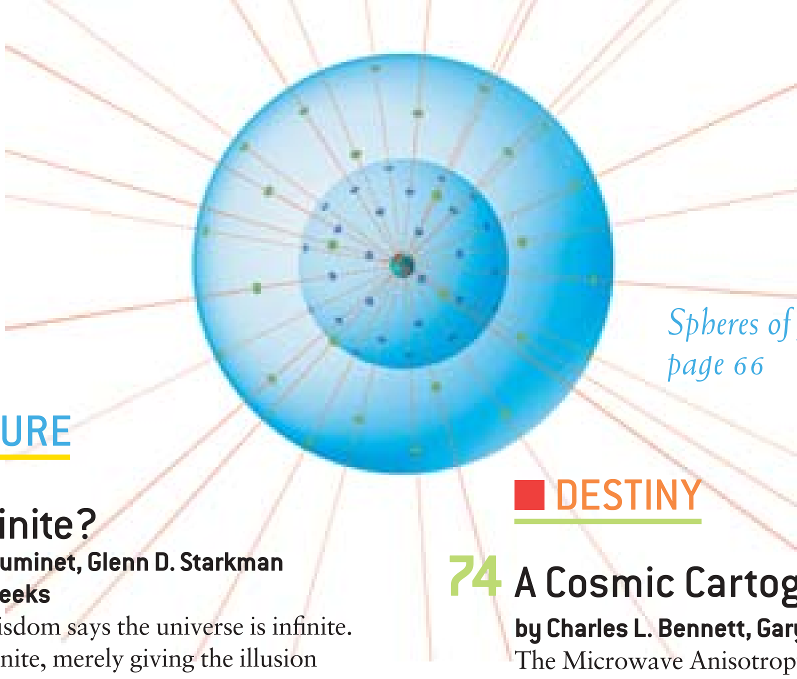
SCIENTIFIC AMERICAN Volume 12 Number 2

INTRODUCTION

EVOLUTION

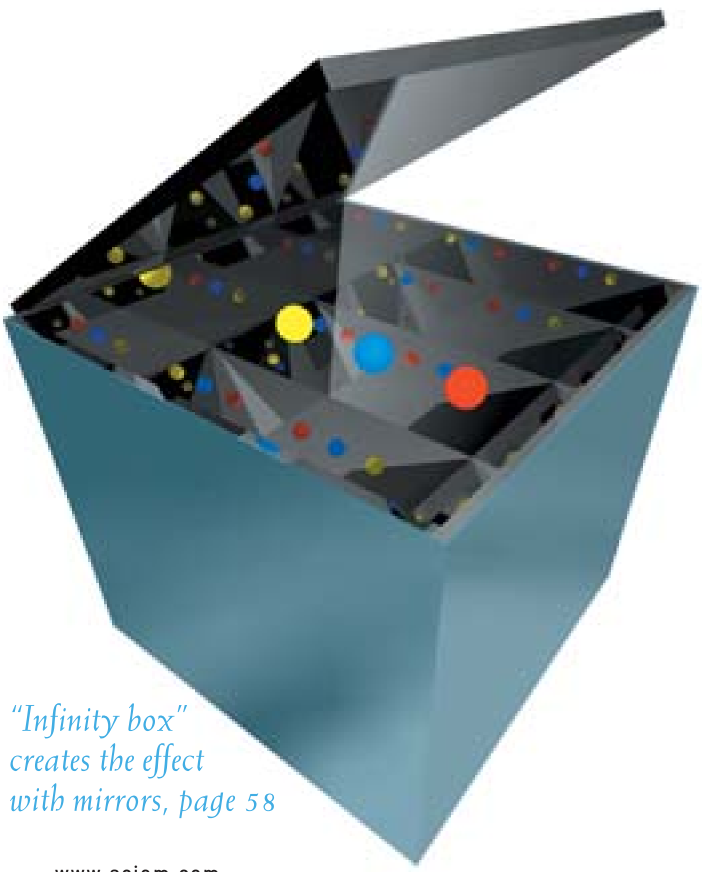
EXPANSION

Cone Nebula, captured in April 2002 by the Hubble Space Telescope



Spheres of gravitational influence, page 66

STRUCTURE



"Infinity box" creates the effect with mirrors, page 58

DESTINY

Cover illustration by Edwin Faughn; NASA/Associated Press (*opposite page*); Bryan Christie Design (*left* and *above*)

Scientific American Special (ISSN 1048-0943), Volume 12, Number 2, 2002, published by Scientific American, Inc., 415 Madison Avenue, New York, NY 10017-1111.

Canadian BN No. 127387652RT; QST No. Q1015332537. To purchase additional quantities: 1 to 9 copies: U.S. $5.95 each plus $2.00 per copy for postage and handling (outside U.S. $5.00 P&H). Send payment to Scientific American, Dept. COSM, 415 Madison Avenue, New York, NY 10017-1111. Inquiries: Fax 212-355-0408 or telephone 212-451-8890. Printed in U.S.A.

Confused by all those theories? Good

making sense of modern cosmology

BY P. JAMES E. PEEBLES

This is an exciting time for cosmologists: findings are pouring in, ideas are bubbling up, and research to test those ideas is simmering away. But it is also a confusing time. All the ideas under discussion cannot possibly be right; they are not even consistent with one another. How is one to judge the progress? Here is how I go about it.

For all the talk of overturned theories, cosmologists have firmly established the foundations of our field. Over the past 70 years we have gathered abundant evidence that our universe is expanding and cooling. First, the light from distant galaxies is shifted toward the red, as it should be if space is expanding and galaxies are pulled away from one another. Second, a sea of thermal radiation fills space, as it should if space used to be denser and hotter. Third, the universe contains large amounts of deuterium and helium, as it should if temperatures were once much higher. Fourth, distant galaxies, seen as they were in the past because of light's travel time, look distinctly younger, as they should if they are closer to the time when no galaxies existed. Finally, the curvature of spacetime seems to be related to the material content of the universe, as it should be if the universe is expanding according to the predictions of Einstein's gravity theory, the general theory of relativity.

That the universe is expanding and cooling is the essence of the big bang theory. You will notice I have said nothing about an "explosion"—the big bang theory describes how our universe is evolving, not how it began.

I compare the process of establishing such compelling results, in cosmology or any other science, to the assembly of a framework. We seek to reinforce each piece of evidence by adding cross bracing from diverse measurements. Our framework for the expansion of the universe is braced tightly enough to be solid. The big bang theory is no longer seriously questioned; it fits together too well. Even the most radical alternative—the latest incarnation of the steady state theory—does not dispute that the universe is expanding and cooling. You still hear differences of opinion in cosmology, to be sure, but they concern additions to the solid part.

For example, we do not know what the universe was doing before it was expanding. A leading theory, inflation, is an attractive addition to the framework, but it lacks cross bracing. That is precisely what cosmologists are now seeking. If mea-

THE AUTHOR

P. JAMES E. PEEBLES is one of the world's most distinguished cosmologists, a key player in the early analysis of the cosmic microwave background radiation and the bulk composition of the universe. He has received some of the highest awards in astronomy, including the 1982 Heineman Prize, the 1993 Henry Norris Russell Lectureship of the American Astronomical Society and the 1995 Bruce Medal of the Astronomical Society of the Pacific. He is emeritus professor at Princeton University.

Updated from the January 2001 issue

REPORT CARD FOR MAJOR THEORIES

Concept	Grade	Comments
The universe evolved from a hotter, denser state	A+	Compelling evidence drawn from many corners of astronomy and physics
The universe expands as the general theory of relativity predicts	A–	Passes the tests so far, but few of the tests have been tight
Dark matter made of exotic particles dominates galaxies	B+	Many lines of indirect evidence, but the particles have yet to be found and alternative theories have yet to be ruled out
Most of the mass of the universe is smoothly distributed; it acts like Einstein's cosmological constant, causing the expansion to accelerate	B–	Encouraging fit from recent measurements, but more must be done to improve the evidence and resolve the theoretical conundrums
The universe grew out of inflation	Inc	Elegant, but lacks direct evidence and requires huge extrapolation of the laws of physics

surements in progress agree with the unique signatures of inflation, then we will count them as a persuasive argument for this theory. But until that time, I would not settle any bets on whether inflation really happened. I am not criticizing the theory; I simply mean that this is brave, pioneering work still to be tested.

More solid is the evidence that most of the mass of the universe consists of dark matter clumped around the outer parts of galaxies. We also have a reasonable case for Einstein's infamous cosmological constant or something similar; it would be the agent of the acceleration that the universe now seems to be undergoing. A decade ago cosmologists generally welcomed dark matter as an elegant way to account for the motions of stars and gas within galaxies. Most researchers, however, had a real distaste for the cosmological constant. Now the majority accept it, or its allied concept, quintessence. Particle physicists have come to welcome the challenge that the cosmological constant poses for quantum theory. This shift in opinion is not a reflection of some inherent weakness; rather it shows the subject in a healthy state of chaos around a slowly growing fixed framework. We students of nature adjust our concepts as the lessons continue.

The lessons, in this case, include the signs that cosmic expansion is accelerating: the brightness of supernovae near and far; the ages of the oldest stars; the bending of light around distant masses; and the fluctuations of the temperature of the thermal radiation across the sky. The evidence is impressive, but I am still skeptical about details of the case for the cosmological constant, including possible contradictions with the evolution of galaxies and their spatial distribution. The theory of the accelerating universe is a work in progress. I admire the architecture, but I would not want to move in just yet.

How might one judge reports in the media on the progress of cosmology? I feel uneasy about articles based on an interview with just one person. Research is a complex and messy business. Even the most experienced scientist finds it hard to keep everything in perspective. How do I know that this individual has managed it well? An entire community of scientists can head off in the wrong direction, too, but it happens less often. That is why I feel better when I can see that the journalist has consulted a cross section of the community and has found agreement that a certain result is worth considering. The result becomes more interesting when others reproduce it. It starts to become convincing when independent lines of evidence point to the same conclusion. To my mind, the best media reports on science describe not only the latest discoveries and ideas but also the essential, if sometimes tedious, process of testing and installing the cross bracing.

Over time, inflation, quintessence and other concepts now under debate either will be solidly integrated into the central framework or will be abandoned and replaced by something better. In a sense, we are working ourselves out of a job. But the universe is a complicated place, to put it mildly, and it is silly to think we will run out of productive lines of research anytime soon. Confusion is a sign that we are doing something right: it is the fertile commotion of a construction site. SA

MORE TO EXPLORE

The Evolution of the Universe. P. James E. Peebles, David N. Schramm, Edwin L. Turner and Richard G. Kron in *Scientific American,* Vol. 271, No. 4, pages 52–57; October 1994.

The Inflationary Universe: The Quest for a New Theory of Cosmic Origins. Alan H. Guth. Perseus Press, 1997.

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

The Accelerating Universe: Infinite Expansion, the Cosmological Constant, and the Beauty of the Cosmos. Mario Livio and Allan Sandage. John Wiley & Sons, 2000.

EVOLUTION

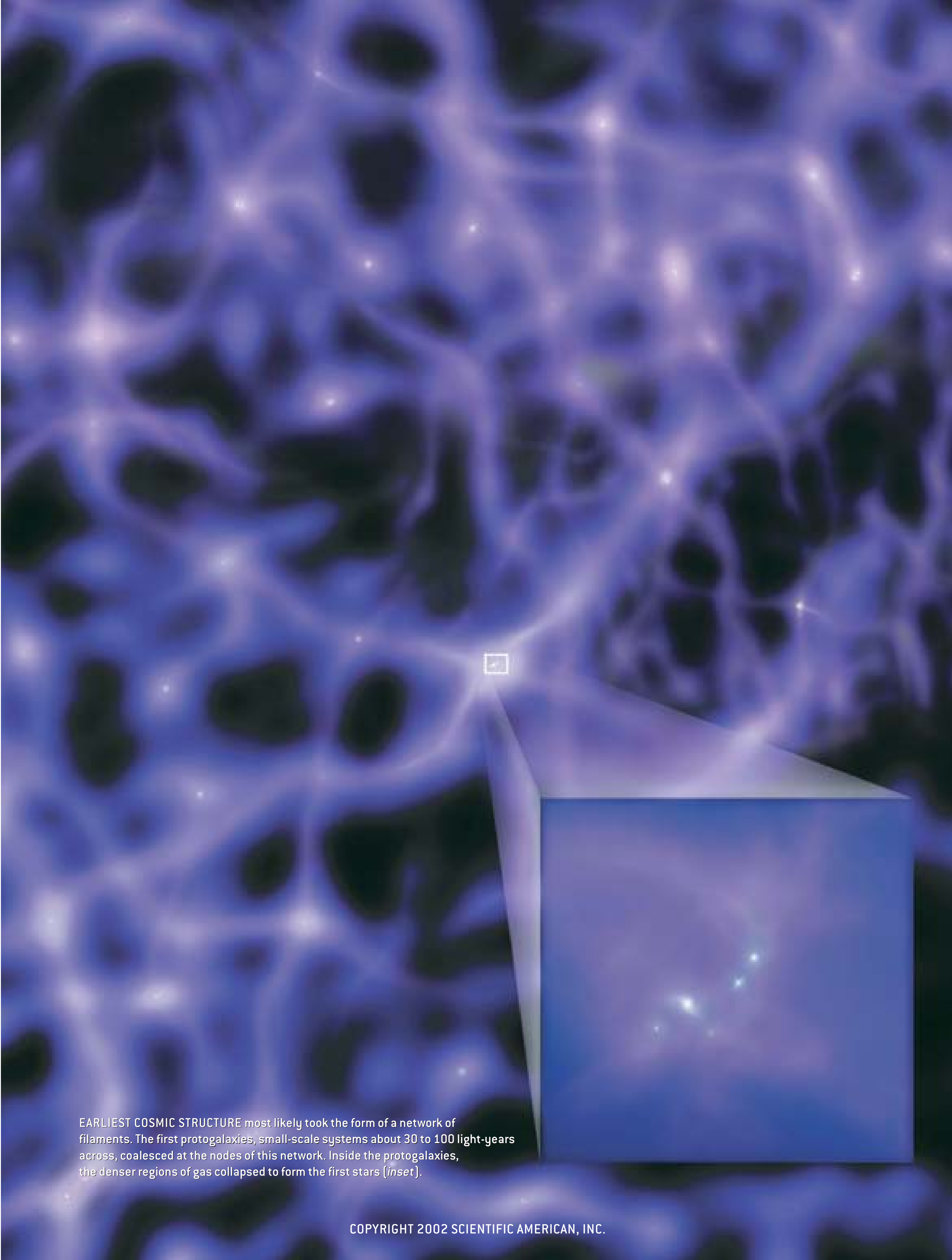
THE FIRST STARS IN THE UNIVERSE

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 about 14 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. The cosmological models predict that 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 cosmological 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 about 30 to 100 light-years across. These properties are similar to those of the molecular gas clouds in which stars are currently forming in the Milky Way, but the first protogalaxies would have differed in some fundamental ways. For one, they would have consisted mostly of dark matter, the putative elementary particles that are believed to make up about 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

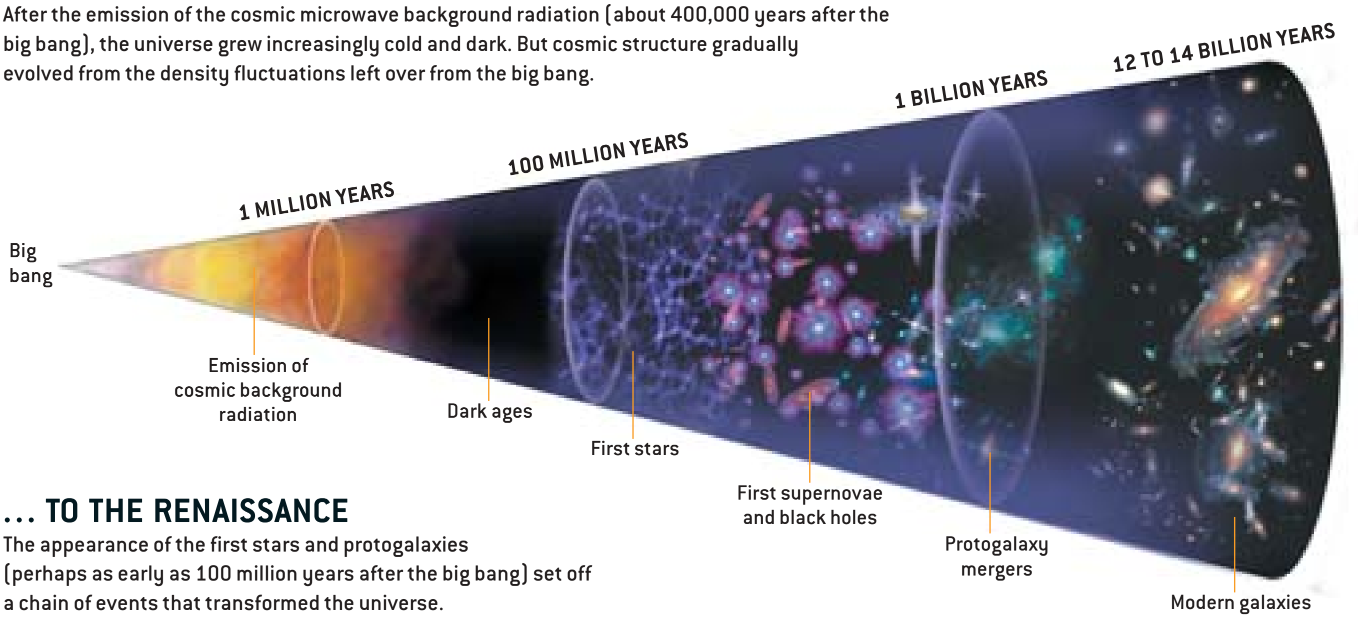
Overview/*The First Stars*

- 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.

COSMIC TIME LINE

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.

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; following this terminology, 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. Therefore, scientists may find it easier to model the formation of the first stars than to model how stars form at present. In any case, the problem is an appealing one for theoretical study, and several research groups have used computer simulations to portray the formation of the earliest stars.

A group consisting of Tom Abel, Greg Bryan and Michael L. Norman (now at Pennsylvania State University, the Massachusetts Institute of Technology 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. Fumitaka Nakamura and Masayuki Umemura (now at Niigata and Tsukuba universities in Japan, respectively) have worked with a more idealized simulation, but it has still yielded instructive results. Although these studies differ in various details, they have all produced similar descriptions of how the earliest stars might have been born.

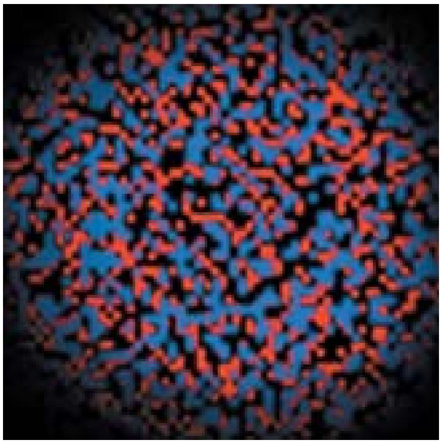
Let There Be Light!

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 about 200 to 300 kelvins, reducing the gas pressure in these regions and hence 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 settles into a flattened rotating configuration that is 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

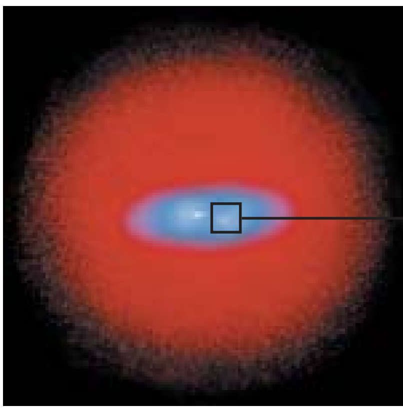
THE BIRTH AND DEATH OF THE FIRST STARS



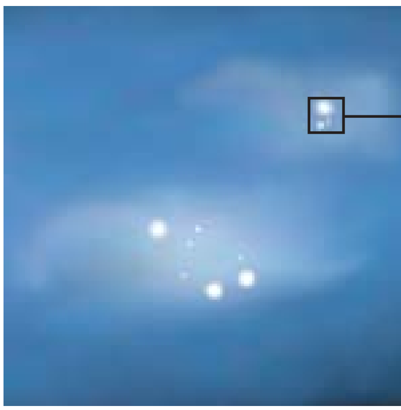
PRIMEVAL TURMOIL

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.

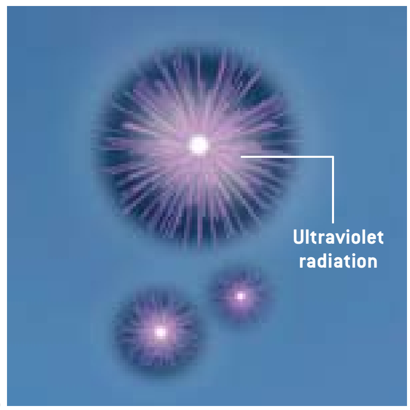
1 The first star-forming systems—small protogalaxies—consisted mostly of the elementary particles known as dark matter (*shown in red*). Ordinary matter—mainly hydrogen gas (*blue*)—was initially mixed with the dark matter.



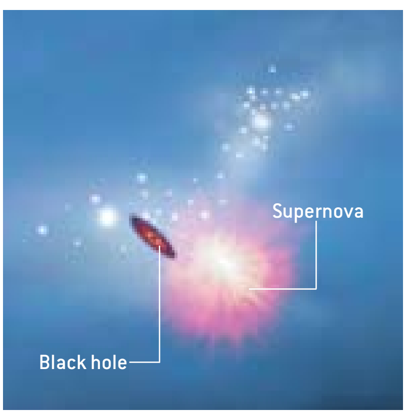
2 The cooling of the hydrogen allowed the ordinary matter to contract, whereas the dark matter remained dispersed. The hydrogen settled into a disk at the center of the protogalaxy.



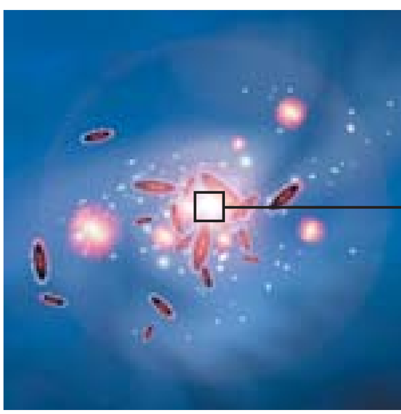
3 The denser regions of gas contracted into star-forming clumps, each hundreds of times as massive as the sun. Some of the clumps of gas collapsed to form very massive, luminous stars.



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.

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 than those of molecular clouds, 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 observed in these clouds are about the same. If we scale up by a factor of almost 1,000, we can estimate that the masses of the first star-forming clumps would have been about 500 to 1,000 solar masses. In agreement with this prediction, all 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 (for example, the exact nature of the initial density fluctuations). In fact, 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 down the contraction until the clumps have built up to at least a few hundred solar masses.

What was the fate of the first collapsing clumps? Did they form stars with similarly large masses, or did they fragment into many smaller parts 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; observations and simulations show that the fragmentation of star-forming 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.

Different groups have arrived at somewhat different estimates of just how massive the first stars might have been. Abel, Bryan and Norman have argued that the stars probably had masses no greater than 300 solar masses. Our own work suggests that masses as high as 1,000 solar masses might have been possible. Both predictions might be valid in different circumstances: the very first stars to form might have had masses no larger than 300 solar masses, whereas stars that formed a little later from the collapse of larger protogalaxies might have reached the higher estimate. Quantitative predictions are difficult because of feedback effects; as a massive star forms, it produces intense radiation and matter outflows that may blow away some of the gas in the collapsing clump. But these effects depend strongly on the presence of heavy elements in the gas, and therefore they should be less important for the earliest stars. Thus, it seems safe to conclude that the first stars in the universe were typically many times more massive and luminous than the sun.

It seems safe to conclude that the **first stars** in the universe 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 Rolf-Peter Kudritzki of the University of Hawaii and Loeb of Harvard, one of us (Bromm) devised theoretical models of such stars with masses between 100 and 1,000 solar masses. The models

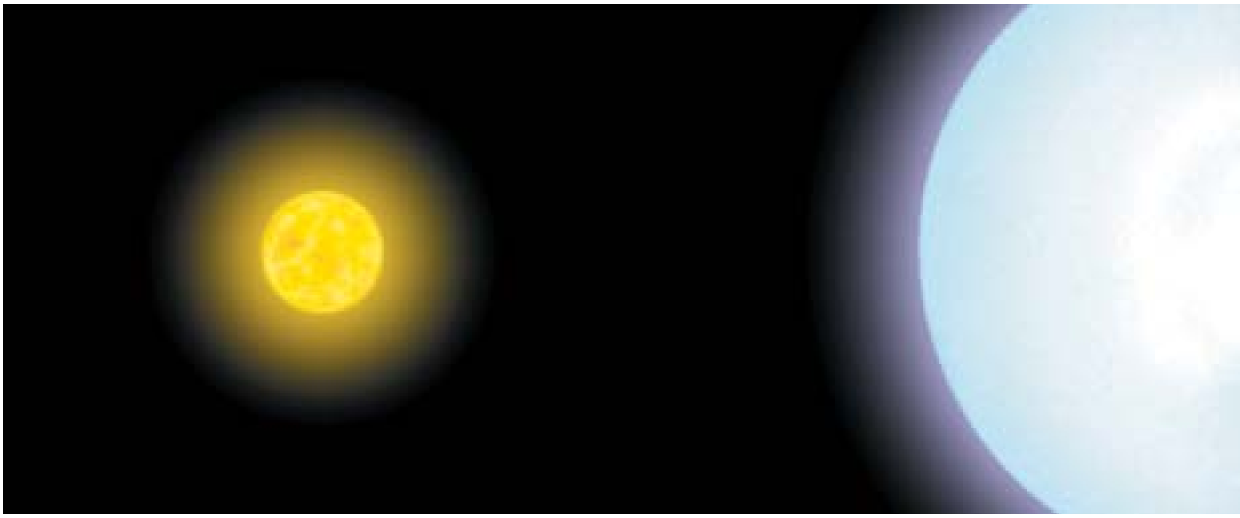
THE AUTHORS

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 a postdoctoral researcher at the Harvard-Smithsonian Center for Astrophysics, where he focuses on the emergence of cosmic structure. The authors acknowledge the many contributions of Paolo Coppi, associate 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**

showed that the stars had surface temperatures of about 100,000 kelvins—about 17 times higher than the sun's surface temperature. Therefore, 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 refer to this event as the cosmic renaissance. Although astronomers cannot yet estimate how much of the gas in the universe condensed into the first stars, even a fraction as small 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 star. As more and more stars began to form over many hundreds of millions of years, the bubbles of ionized gas would have eventually merged, and the intergalactic gas would have become completely ionized.

Scientists from the California Institute of Technology and the Sloan Digital Sky Survey have recently found evidence for the final stages of this ionization process. The researchers 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. 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 at the end of their lives, expelling the metals they produced by fusion reactions. 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 University of Florence 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 a 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 overall rate at which stars are born. In fact, 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 the temperature 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, large numbers of 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. Such events are now fairly rare, but some evidence suggests that they may produce relatively large numbers of massive stars.

Second-generation stars might have been primarily responsible for the **cosmic renaissance**.

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. The case for a high supernova rate at early times 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 similarly massive black holes. Several of the computer 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—millions of times more massive than the sun—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 Next Generation 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

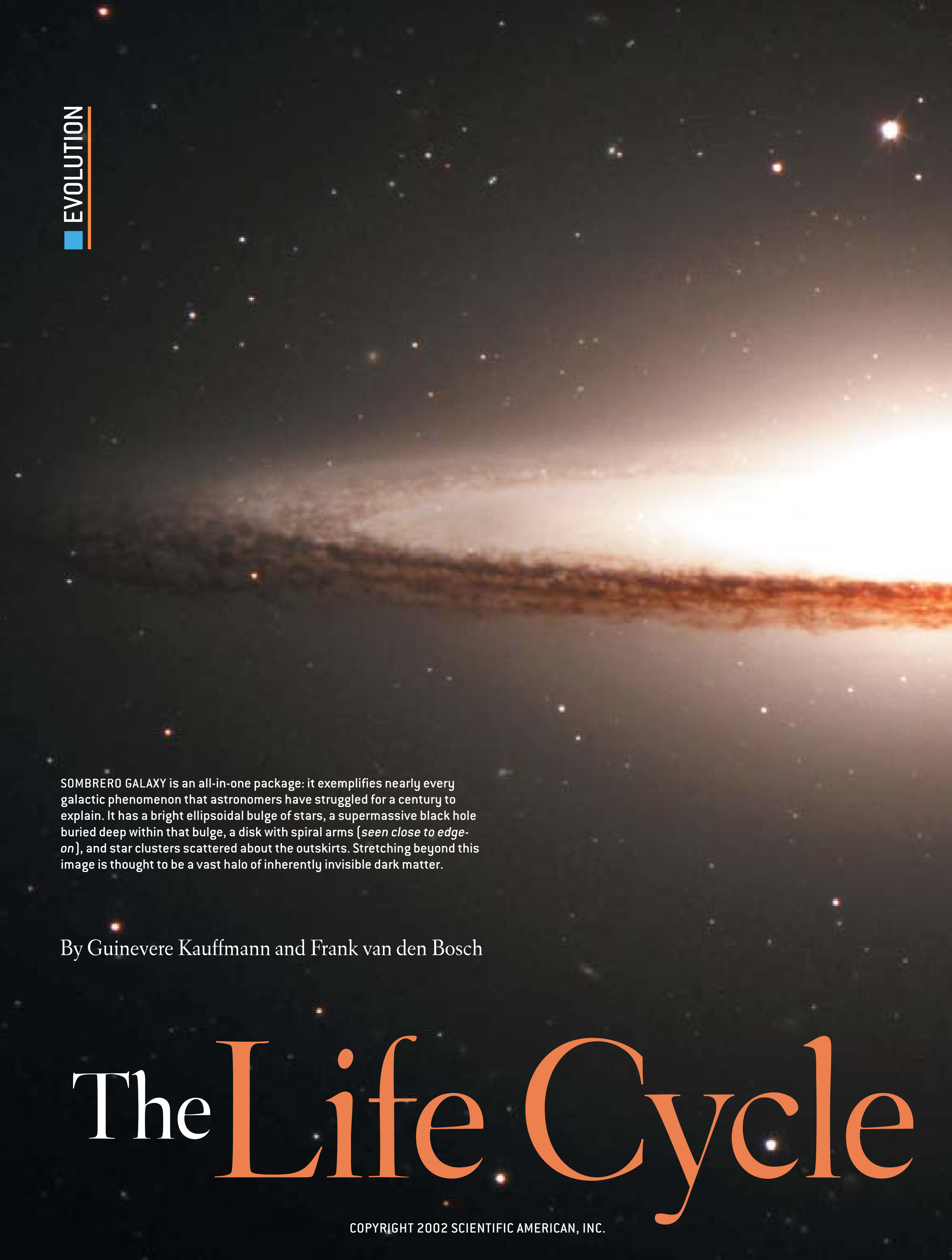
MORE TO EXPLORE

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

The Formation of the First Stars. Richard B. Larson in *Star Formation from the Small to the Large Scale.* Edited by F. Favata, A. A. Kaas and A. Wilson. ESA Publications, 2000. Available on the Web at **www.astro.yale.edu/larson/papers/Noordwijk99.pdf**

In the Beginning: The First Sources of Light and the Reionization of the Universe. R. Barkana and A. Loeb in *Physics Reports,* Vol. 349, No.2, pages 125–238; July 2001. Available on the Web at **aps.arxiv.org/abs/astro-ph/0010468**

Graphics from computer simulations of the formation of the first stars can be found at **www.tomabel.com**

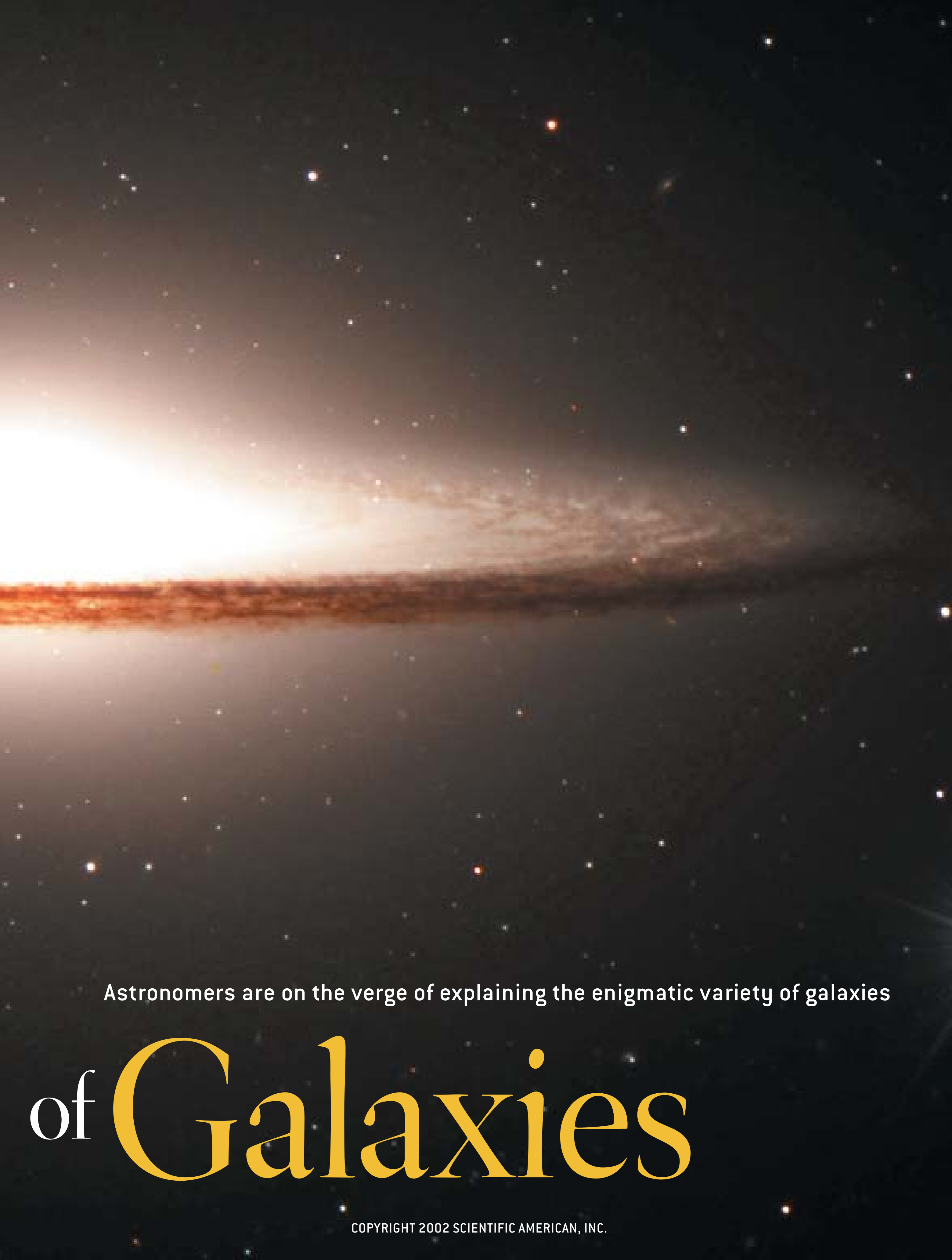


EVOLUTION

SOMBRERO GALAXY is an all-in-one package: it exemplifies nearly every galactic phenomenon that astronomers have struggled for a century to explain. It has a bright ellipsoidal bulge of stars, a supermassive black hole buried deep within that bulge, a disk with spiral arms (*seen close to edge-on*), and star clusters scattered about the outskirts. Stretching beyond this image is thought to be a vast halo of inherently invisible dark matter.

By Guinevere Kauffmann and Frank van den Bosch

The Life Cycle



Astronomers are on the verge of explaining the enigmatic variety of galaxies

# of Galaxies

In many science-fiction stories,

a mighty empire dooms itself through its hubris: it presumes to conquer and rule an entire galaxy. That seems a lofty ambition indeed. To bring our Milky Way galaxy to heel, an empire would have to vanquish 100 billion stars. But cosmologists—those astronomers who study the universe as a whole—are unimpressed. The Milky Way is one of 50 billion or more galaxies within the observable reaches of space. To conquer it would be to conquer an insignificant speck.

A century ago nobody knew all those galaxies even existed. Most astronomers thought that the galaxy and the universe were synonymous. Space contained perhaps a billion stars, interspersed with fuzzy splotches that looked like stars in the process of forming or dying. Then, in the early decades of the 20th century, came the golden age of astronomy, when American astronomer Edwin Hubble and others determined that those fuzzy splotches were often entire galaxies in their own right.

Why do stars reside in gigantic agglomerations separated by vast voids, and how do galaxies take on their bewildering variety of shapes, sizes and masses? These questions have consumed astronomers for decades. It is not possible for us to observe a galaxy forming; the process is far too slow. Instead researchers have to piece the puzzle together by observing many different galaxies, each caught at a different phase in its evolutionary history. Such measurements did not become routine until about a decade ago, when astronomy entered a new golden age.

Spectacular advances in telescope and detector technology are now giving astronomers a view of how galaxies have changed over cosmic timescales. The Hubble Space Telescope has taken very deep snapshots of the sky, revealing galaxies down to unprecedentedly faint levels. Ground-based instruments such as the giant Keck telescopes have amassed statistics on distant (and therefore ancient) galaxies. It is as if evolutionary biologists had been handed a time machine, allowing them to travel back into prehistory and take pictures of the animals and plants inhabiting the earth at a series of different epochs. The challenge for astronomers, as it would be for the biologists, is to determine how the species observed at the earliest times evolved into what we know today.

The task is of truly astronomical proportions. It involves physics on wildly disparate scales, from the cosmological evolution of the entire universe to the formation of a single star. That makes it difficult to build realistic models of galaxy formation, yet it brings the whole subject full circle. The discovery of all those billions of galaxies made stellar astronomy and cosmology seem mutually irrelevant. In the grand scheme of things, stars were just too small to matter; conversely, debates over the origin of the universe struck most stellar astronomers as hopelessly abstract. Now we know that a coherent picture of the universe must take in both the large and the small.

Overview/*Galaxy Evolution*

- One of the liveliest subfields of astrophysics right now is the study of how galaxies take shape. Telescopes are probing the very earliest galaxies, and computer simulations can track events in unprecedented detail.
- Researchers may soon do for galaxies what they did for stars in the early 20th century: provide a unified explanation, based on a few general processes, for a huge diversity of celestial bodies. For galaxies, those processes include gravitational instability, radiative cooling and star formation, relaxation (galaxies reach internal equilibrium) and interactions among galaxies.
- Several vexing questions remain, however. A possible answer to these questions is that supernova explosions actually have a profound and pervasive effect on their structure.

Galactic Species

TO UNDERSTAND HOW galaxies form, astronomers look for patterns and trends in their properties. According to the classification scheme developed by Hubble, galaxies may be broadly divided into three major types: elliptical, spiral and irregular [*see illustration on opposite page*]. The most massive ones are the ellipticals. These are smooth, featureless, almost spherical systems with little or no gas or dust. In them, stars buzz around the center like bees around a hive. Most of the stars are very old.

Spiral galaxies, such as our own Milky Way, are highly flattened and organized structures in which stars and gas move on circular or near-circular orbits around the center. In fact, they are also known as disk galaxies. The pinwheel-like spiral arms are filaments of hot young stars, gas and dust. At their centers, spiral galaxies contain bulges—spheroidal clumps of stars that are reminiscent of miniature elliptical galaxies. Roughly a third of spiral galaxies have a rectangular structure toward the cen-

EUROPEAN SOUTHERN OBSERVATORY/BARTHEL/NEESER (*preceding pages*)

TYPES OF GALAXIES

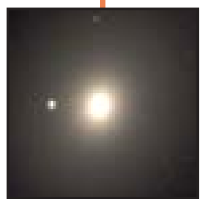
ASTRONOMERS SORT GALAXIES using the "tuning fork" classification scheme developed by American astronomer Edwin Hubble in the 1920s. According to this system, galaxies come in three basic types: elliptical (*represented by the handle of the fork at right*), spiral (*shown as prongs*) and irregular (*shown below at left*). The smallest galaxies, known as dwarfs, have their own uncertain taxonomy.

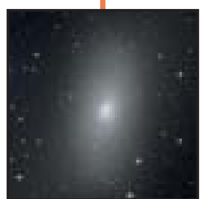
Within each of the types are subtypes that depend on the details of the galaxy's shape. Going from the top of the tuning fork to the bottom, the galactic disk becomes more prominent in optical images and the central bulge less so. The different Hubble types may represent various stages of development. Galaxies start off as spirals without bulges, undergo a collision during which they appear irregular, and end up as ellipticals or as spirals with bulges.

—*G.K.* and *F.v.d.B.*

ELLIPTICALS


M89
E0


M49
E4


M110
E5

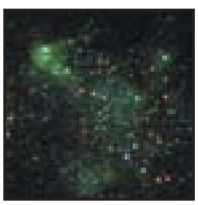

M84
S0

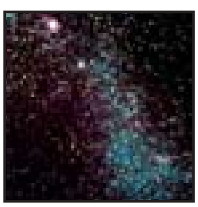
IRREGULARS


M82
Irregular

DWARF TYPES


M32
Elliptical

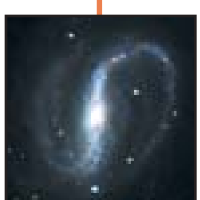

VII Zw 403
Blue Compact

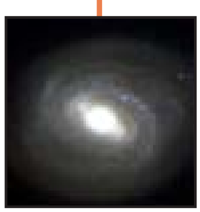

Small Magellanic Cloud
Irregular


Leo I
Spheroidal

BARRED SPIRALS


NGC 660
SBa

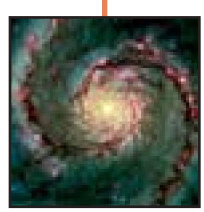

NGC 7479
SBb


M58
SBc

NORMAL SPIRALS

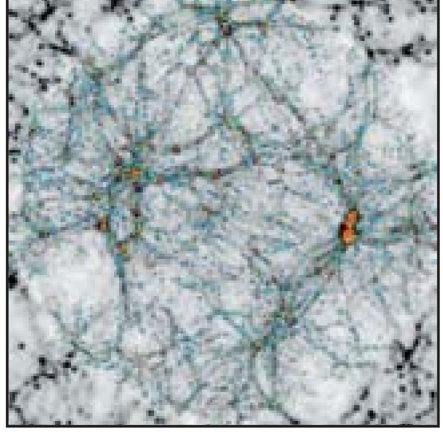

NGC 7217
Sa


NGC 4622
Sb


M51
Sc

N. A. SHARP/NOAO/AURA/NSF (*M82*); B. KEEL/HALL TELESCOPE/LOWELL OBSERVATORY (*M32*); R. SCHULTE-LADBECK/U. HOPP/M. CRONE/*ASTROPHYSICAL JOURNAL* (*blue compact dwarf*); NOAO/AURA/NSF (*Small Magellanic Cloud*); DAVID MALIN, © ANGLO-AMERICAN OBSERVATORY (*Leo I*); NOAO/AURA/NSF (*M89, M49, M110, M84*); R. BRANCH/R. MILNER/A. BLOCK/NOAO/AURA/NSF (*NGC 660*); A. BLOCK/NOAO/AURA/NSF (*NGC 7479*); F. CIESLAK/A. BLOCK/NOAO/AURA/NSF (*M58*); B. KEEL/R. BUTA/G. PURCELL/CERRO TOLOLO INTER-AMERICAN OBSERVATORY, CHILE (*NGC 7217*); G. BYRD/R. BUTA/T. FREEMAN/NASA (*NGC 4622*); NASA/STSCI/AURA (*M51*)

Supercomputer simulations of the spatial distribution of galaxies are in excellent agreement with observations.



H. MATHIS, V. SPRINGEL, G. KAUFFMANN AND S.D.M. WHITE *Max Planck Institute for Astrophysics, Garching, Germany,* AND G. LEMSON, A. ELDAR AND A. DEKEL *Hebrew University, Israel (simulation of galaxy formation in a region 900 million light-years across)*

ter. Such "bars" are thought to arise from instabilities in the disk.

Irregular galaxies are those that do not fit into the spiral or elliptical classifications. Some appear to be spirals or ellipticals that have been violently distorted by a recent encounter with a neighbor. Others are isolated systems that have an amorphous structure and exhibit no signs of any recent disturbance.

Each of these three classes covers galaxies with a wide range of luminosities. On average, however, ellipticals are brighter than spirals, and fainter galaxies are more likely than their luminous counterparts to be irregular. For the faintest galaxies, the classification scheme breaks down altogether. These dwarf galaxies are heterogeneous in nature, and attempts to pigeonhole them have proved controversial. Loosely speaking, they fall into two categories: gas-rich systems where stars are actively forming and gas-poor systems where no stars are forming.

An important clue to the origin of the galaxy types comes from the striking correlation between type and local galaxy density. Most galaxies are scattered through space far from their nearest neighbor, and of these only 10 to 20 percent are ellipticals; spirals dominate. The remaining galaxies, however, are packed into clusters, and for them the situation is reversed. Ellipticals are the majority, and the spirals that do exist are anemic systems depleted of gas and young stars. This so-called morphology-density relation has long puzzled astronomers.

Light and Dark

A SMALL PERCENTAGE of spirals and ellipticals are peculiar in that they contain an exceedingly luminous, pointlike core—an active galactic nucleus (AGN). The most extreme and rarest examples are the quasars, which are so bright that they completely outshine their host galaxies. Astronomers generally believe that AGNs are powered by black holes weighing millions to billions of solar masses. Theory predicts that gas falling into these monsters will radiate about 10 percent of its intrinsic energy, sufficient to generate a beacon that can be detected on the other side of the universe.

Once considered anomalies, AGNs have recently been shown to be integral to the process of galaxy formation. The peak of AGN activity occurred when the universe was approximately a fourth of its present age—the same time that most of the stars in ellipticals were being formed. Furthermore, supermassive black holes are now believed to reside in virtually every elliptical galaxy, as well as every spiral galaxy that has a bulge, regardless of whether those galaxies contain an AGN [see "The Hole Shebang," by George Musser; News and Analysis, Scientific American, October 2000]. The implication is that every galaxy may go through one or more episodes of AGN activity. As long as matter falls into the black hole, the nucleus is active. When no new material is supplied to the center, it lies dormant.

Most of the information we have about all these phenomena comes from photons: optical photons from stars, radio photons from neutral hydrogen gas, x-ray photons from ionized gas. But the vast majority of the matter in the universe may not emit photons of any wavelength. This is the infamous dark matter, whose existence is inferred solely from its gravitational effects. The visible parts of galaxies are believed to be enveloped in giant "halos" of dark matter. These halos, unlike those found above the heads of saints, have a spherical or ellipsoidal shape. On larger scales, analogous halos are thought to keep clusters of galaxies bound together.

Unfortunately, no one has ever detected dark matter directly, and its nature is still one of the biggest mysteries in science. Currently most astronomers favor the idea that dark matter consists mostly of hitherto unidentified particles that barely interact with ordinary particles or with one another. Astronomers typically refer to this class of particles as cold dark matter (CDM) and any cosmological model that postulates their existence as a CDM model.

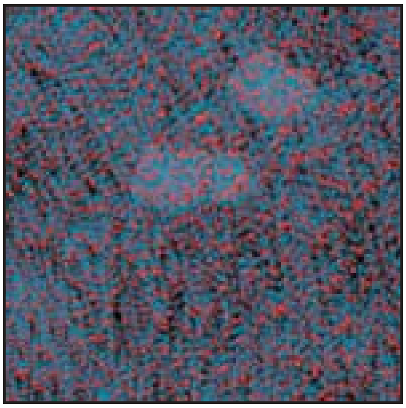
Over the past two decades, astronomers have painstakingly developed a model of galaxy formation based on CDM. The basic framework is the standard big bang theory for the expansion of the universe. Cosmologists continue to debate how the expansion got going and what transpired early on, but these uncertainties do not matter greatly for galaxy formation. We pick up the story about 100,000 years after the big bang, when the universe consisted of baryons (that is, ordinary matter, predominantly hydrogen and helium nuclei), electrons (bound to the nuclei), neutrinos, photons and CDM. Observations indicate that the matter and radiation were distributed smoothly:

THE AUTHORS

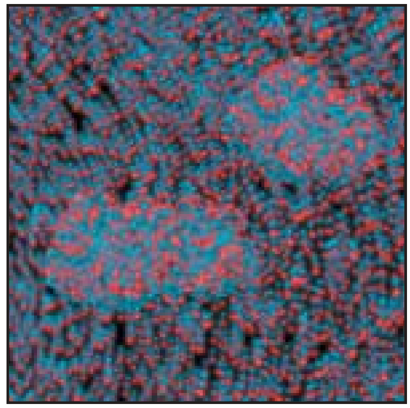
GUINEVERE KAUFFMANN and *FRANK VAN DEN BOSCH* are researchers at the Max Planck Institute for Astrophysics in Garching, Germany. They are among the world's experts on the theoretical modeling of galaxy formation. Kauffmann has recently turned her attention to analyzing data from the Sloan Digital Sky Survey, which she believes holds the answers to some of the mysteries highlighted in this article. In her spare time, she enjoys exploring Bavaria with her son, Jonathan. Van den Bosch is particularly intrigued by the formation of disk galaxies and of massive black holes in galactic centers. In his free time, he can often be found in a Munich beer garden.

COOKING UP A GALAXY

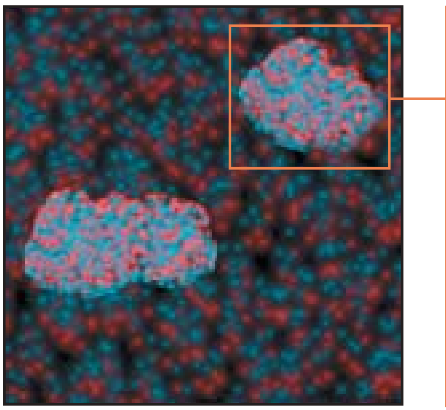
THREE BASIC PROCESSES dictated how the primordial soup congealed into galaxies: the overall expansion of the universe in the big bang, the force of gravity, and the motion of particles and larger constituents. The shifting balance among these processes can explain why galaxies became discrete, coherent bodies rather than a uniform gas or a horde of black holes. In this theory, small bodies coalesce first and then glom together to form larger objects. A crucial ingredient is dark matter, which reaches a different equilibrium than ordinary matter. *—G.K. and F.v.d.B.*



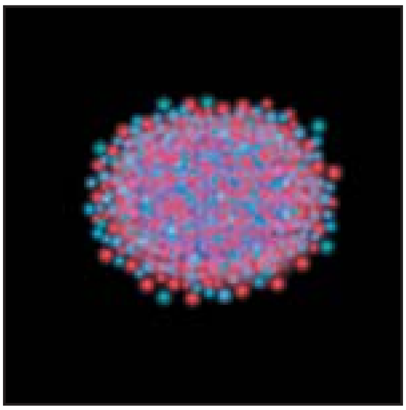
1 In the beginning, a primordial fluid—a mixture of ordinary matter (*blue*) and dark matter (*red*)—fills the universe. Its density varies subtly from place to place.



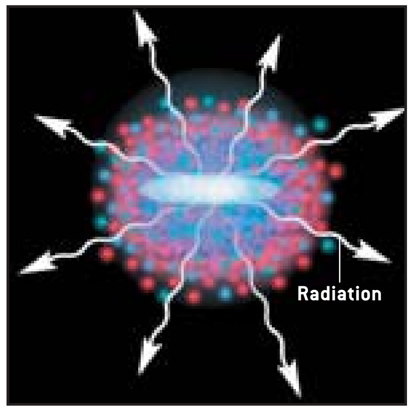
2 At first, cosmic expansion overpowers gravity. The fluid thins out. But patches of higher density thin out more slowly than other regions do.



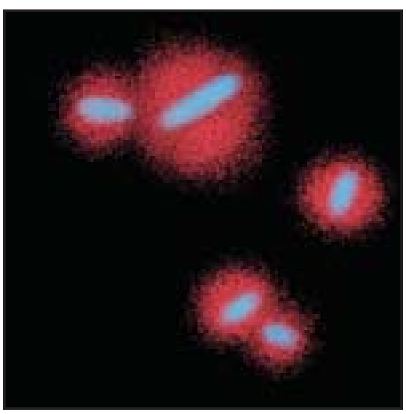
3 Eventually these patches become so dense, relative to their surroundings, that gravity takes over from expansion. The patches start to collapse.



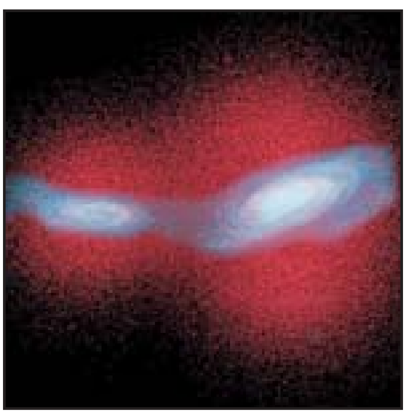
4 As each patch collapses, it attains equilibrium. The density, both of ordinary and of dark matter, peaks at the center and decreases toward the edge.



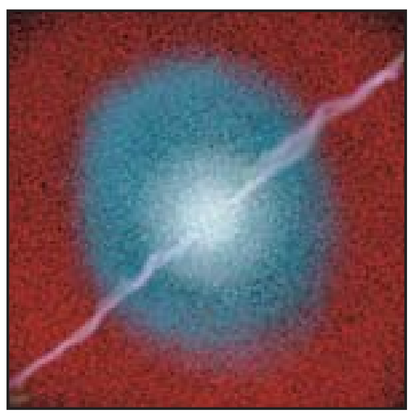
5 Dark matter, being unable to radiate, retains this shape. But ordinary matter emits radiation, collapses into a rotating disk and begins to condense into stars.



6 Protogalaxies interact, exerting torques on one another and merging to form larger and larger bodies. (This step overlaps with steps 4 and 5.)



7 When two disks of similar size merge, the stellar orbits become scrambled. An elliptical galaxy results. Later a disk may develop around the elliptical.



8 The merger triggers new star formation and feeds material into the central black hole, generating an active galactic nucleus, which can spew plasma jets.

DON DIXON

the density at different positions varied by only about one part in 100,000. The challenge is to trace how these simple ingredients could give rise to the dazzling variety of galaxies.

If one compares the conditions back then with the distribution of matter today, two important differences stand out. First, the present-day universe spans an enormous range of densities. The central regions of galaxies are more than 100 billion times as dense as the universe on average. The earth is another 10 billion billion times as dense as that. Second, whereas the baryons and CDM were initially well mixed, the baryons today form dense knots (the galaxies) inside gargantuan halos of dark matter. Somehow the baryons have decoupled from the CDM.

The first of these differences can be explained by the process of gravitational instability. If a region is even slightly more dense than average, the excess mass will exert a slightly stronger-than-average gravitational force, pulling extra matter toward itself. This creates an even stronger gravitational field, pulling in even more mass. This runaway process amplifies the initial density differences.

Sit Back and Relax

ALL THE WHILE, the gravity of the region must compete with the expansion of the universe, which pulls matter apart. Initially cosmic expansion wins and the density of the region decreases. The key is that it decreases more slowly than the density of its surroundings. At a certain point, the overdensity of the region compared with its surroundings becomes so pronounced that its gravitational attraction overcomes the cosmic expansion. The region starts to collapse.

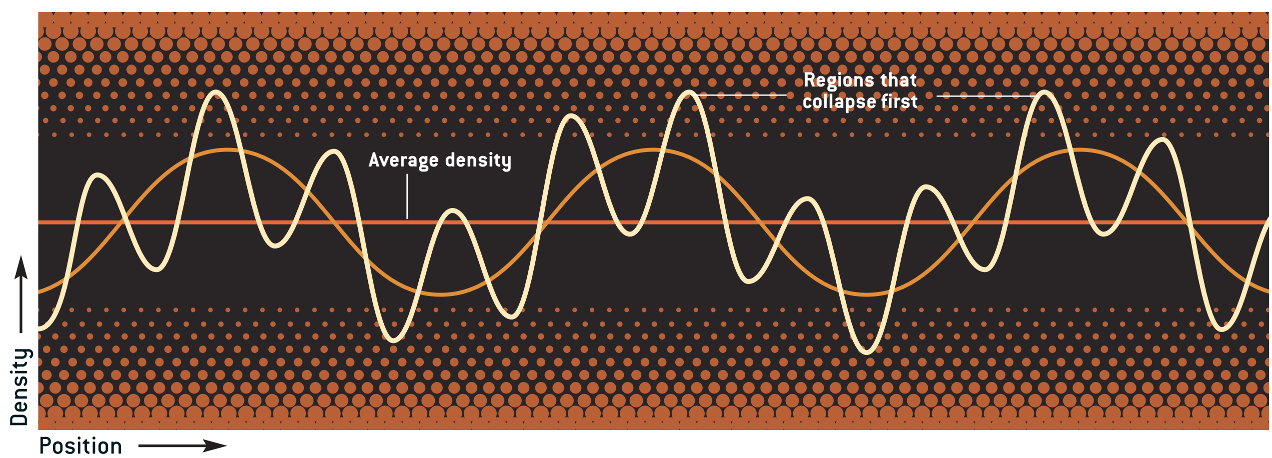
Up to this point, the region is not a coherent object but merely a random enhancement of density in the haze of matter that fills the universe. But once the region collapses, it starts to take on an internal life of its own. The system—which we shall call a protogalaxy from here on—seeks to establish some form of equilibrium. Astronomers refer to this process as relaxation. The baryons behave like the particles of any gas. Heated by shock waves that are triggered by the collapse, they exchange energy through direct collisions with one another, thus achieving hydrostatic equilibrium—a state of balance between pressure and gravity. The earth's atmosphere is also in hydrostatic equilibrium (or nearly so), which is why the pressure decreases exponentially with altitude.

For the dark matter, however, relaxation is distinctively different. CDM particles are, by definition, weakly interactive; they are not able to redistribute energy among themselves by direct collisions. A system of such particles cannot reach hydrostatic equilibrium. Instead it undergoes what is called, perhaps oxymoronically, violent relaxation. Each particle exchanges energy not with another individual particle but with the collective mass of particles, by way of the gravitational field.

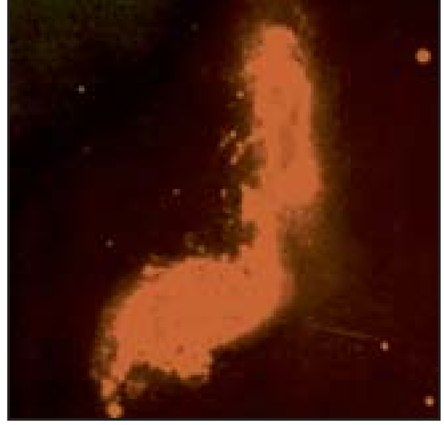
Bodies traveling in a gravitational field are always undergoing an exchange of gravitational and kinetic energy. If you throw a ball into the air, it rises to a higher altitude but decelerates: it gains gravitational energy at the expense of kinetic energy. On the way down, the ball gains kinetic energy at the expense of gravitational energy. CDM particles in a protogalaxy behave much the same way. They move around and change speed as their balance of gravitational and kinetic energy shifts. But unlike balls near the earth's surface, CDM particles move in a gravitational field that is not constant. After all, the gravitational field is produced by all the particles together, which are undergoing collapse.

GALACTIC DENSITY VARIATIONS

DENSITY VARIATIONS in the pregalactic universe followed a pattern that facilitated the formation of protogalaxies. The variations were composed of waves of various wavelengths. A small wave was superimposed on a slightly larger wave, which was superimposed on an even larger wave, and so on. Therefore, the highest density occurred over the smallest regions. These regions collapsed first and became the building blocks for larger structures. —*G.K. and F.v.d.B.*



SARA CHEN



Astronomers may be directly observing, for the first time, the formation of elliptical galaxies.

Changes in the gravitational field cause some particles to gain energy and others to lose energy. Just as for the baryons, this redistribution of the energies of the particles allows the system to relax, forming a CDM halo that is said to be in virial equilibrium. The process is complicated and has never been worked out in great theoretical detail. Instead researchers track it using numerical simulations, which show that all CDM halos in virial equilibrium have similar density profiles.

The end point of the collapse and relaxation of a protogalaxy is a dark matter halo, inside of which the baryonic gas is in hydrostatic equilibrium at a temperature of typically a few million degrees. Whereas each CDM particle conserves its energy from then on, the baryonic gas is able to emit radiation. It cools, contracts and accumulates at the center of the dark matter halo. Cooling, therefore, is the process responsible for decoupling the baryons from the CDM.

So far we have focused on a single protogalaxy and ignored its surroundings. In reality, other protogalaxies will form nearby. Gravity will pull them together until they merge to form a grander structure. This structure will itself merge, and so on. Hierarchical buildup is a characteristic feature of CDM models. The reason is simple. Because small-scale fluctuations in density are superimposed on larger-scale fluctuations, the density reaches its highest value over the smallest regions. An analogy is the summit of a mountain. The exact position of the peak corresponds to a tiny structure: for example, a pebble on top of a rock on top of a hill on top of the summit. If a cloud bank descends on the mountain, the pebble vanishes first, followed by the rock, the hill and eventually the whole mountain.

Similarly, the densest regions of the early universe are the smallest protogalaxies. They are the first regions to collapse, followed by progressively larger structures. What distinguishes CDM from other possible types of dark matter is that it has density fluctuations on all scales. Neutrinos, for example, lack fluctuations on small scales. A neutrino-dominated universe would be like a mountain with an utterly smooth summit.

The hierarchical formation of dark matter halos cannot be described using simple mathematical relationships. It is best studied using numerical simulations. To emulate a representative part of the universe with enough resolution to see the formation of individual halos, researchers must use the latest supercomputers. The statistical properties and spatial distribution of the halos emerging from these simulations are in excellent agreement with those of observed galaxies, providing strong support for the hierarchical picture and hence for the existence of CDM.

Take a Spin

THE HIERARCHICAL PICTURE naturally explains the shapes of galaxies. In spiral galaxies, stars and gas move on circular orbits. The structure of these galaxies is therefore governed by angular momentum. Where does this angular momentum come from? According to the standard picture, when protogalaxies filled the universe, they exerted tidal forces on one another, causing them to spin. After the protogalaxies collapsed, each was left with a net amount of angular momentum.

When the gas in the protogalaxies then started to cool, it contracted and started to fall toward the center. Just as ice-skaters spin faster when they pull in their arms, the gas rotated faster and faster as it contracted. The gas thus flattened out, in the same way that the earth is slightly flatter than a perfect sphere because of its rotation. Eventually the gas was spinning so fast that the centrifugal force (directed outward) became equal to the gravitational pull (directed inward). By the time the gas attained centrifugal equilibrium, it had flattened into a thin disk. The disk was sufficiently dense that the gas started to clump into the clouds, out of which stars then formed. A spiral galaxy was born.

Because most dark matter halos end up with some angular momentum, one has to wonder why all galaxies aren't spirals. How did ellipticals come into being? Astronomers have long held two competing views. One is that most of the stars in present-day ellipticals and bulges formed during a monolithic collapse at early epochs. The other is that ellipticals are relative latecomers, having been produced as a result of the merging of spiral galaxies.

The second view has come to enjoy increasing popularity. Detailed computer simulations of the merger of two spirals show that the strongly fluctuating gravitational field destroys the two disks. The stars within the galaxies are too spread out to bang into one another, so the merging process is quite similar to the violent relaxation suffered by dark matter. If the galaxies are of comparable mass, the result is a smooth clump of stars with properties that strongly resemble an elliptical. Much of the gas in the two original disk galaxies loses its angular momentum and plummets toward the center. There the gas reaches high densities and starts to form stars at a frenzied rate. At later times, new gas may fall in, cool off and build up a new disk around the elliptical. The result will be a spiral galaxy with a bulge in the middle.

The high efficiency of star formation during mergers explains why ellipticals typically lack gas: they have used it up. The merger model also accounts for the morphology-density relation: a galaxy in a high-density environment will undergo more mergers and is thus more likely to become an elliptical.

Observational evidence confirms that mergers and inter-

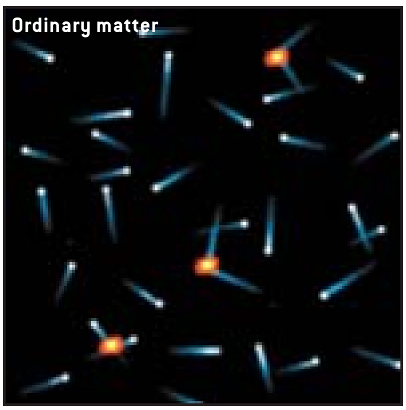
FROM *ATLAS OF PECULIAR GALAXIES*, BY HALTON ARP (*California Institute of Technology, 1966*); nedwww.ipac.caltech.edu/level5/Arp/frames.html (*merging galaxies known as Arp 270*) actions have been common in the universe, particularly early on. In Hubble Space Telescope images, many ancient galaxies have disturbed morphologies, a telltale sign of interaction. Moreover, the number of starburst galaxies—in which stars form at a frenetic pace—increases dramatically at earlier times. Astronomers may be directly observing, for the first time, the formation of elliptical galaxies.

If elliptical galaxies and spiral bulges are linked to galaxy mergers, then it follows that supermassive black holes may be created in these events, too. Hole masses are strongly correlated with the mass of the surrounding elliptical galaxy or bulge; they are not correlated with the mass of the spiral disk. Merger models have been extended to incorporate supermassive holes and therefore AGNs. The abundant gas that is funneled toward the center during a merger could revive a dormant black hole. In other words, quasars were more common in the past because mergers were much more common then.

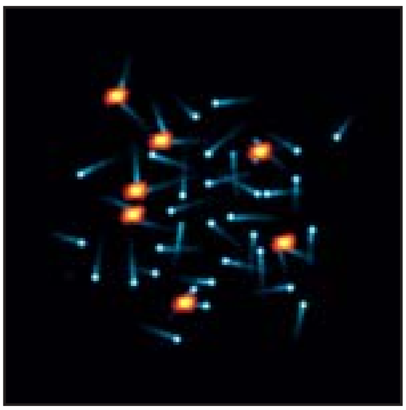
As for dwarf galaxies, in the hierarchical picture they are the leftovers—small clumps that have yet to merge. Recent observations show that star formation in dwarfs is particularly erratic, coming in short bursts separated by long quiescent periods [see "Dwarf Galaxies and Starbursts," by Sara C. Beck; Scientific American, June 2000]. In heftier galaxies such as the Milky Way, star formation occurs at a more constant rate. These results are intriguing because astronomers have often hypothesized that the mass of a galaxy determines its fertility. In lightweight galaxies, supernova explosions can easily disrupt or even rid the system of its gas, thus choking off star formation. Even the smallest perturbation can have a dramatic effect. It is this sensitivity to initial conditions and random events

HOW RELAXING

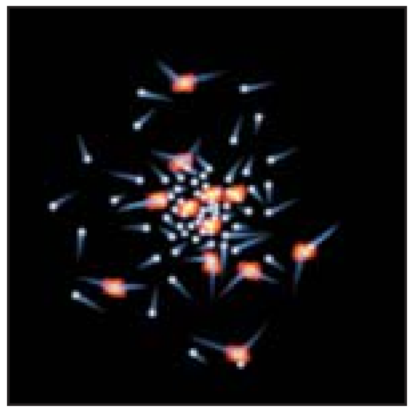
GRAVITY CAUSES small density perturbations to grow until they finally start to collapse. During the collapse the gas and dark matter seek to establish an internal state of equilibrium. This equilibrium determines the overall properties of the galaxy, such as its shape and density profile. The ordinary matter and dark matter attain equilibrium by different means. *—G.K. and F.v.d.B.*



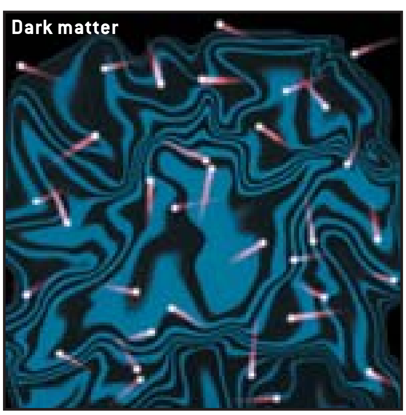
1 The ordinary matter—predominantly hydrogen gas—starts off moving every which way. Its density varies randomly.



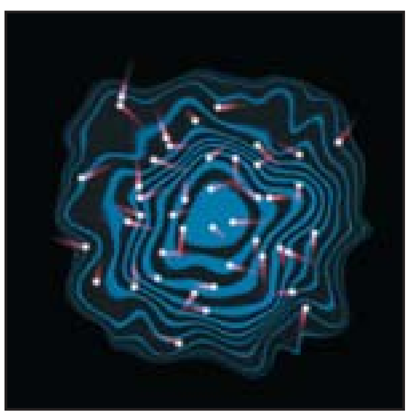
2 The gas particles bang into one another, redistributing energy and generating a pressure that resists gravity.



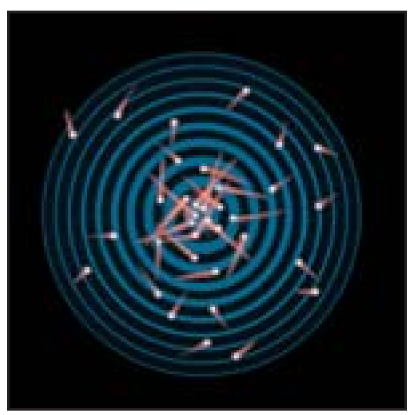
3 Eventually the gas settles down into hydrostatic equilibrium, with the density highest near the center of gravity.



1 Initially the dark matter has the same arrangement as ordinary matter. The difference is that particles do not collide.



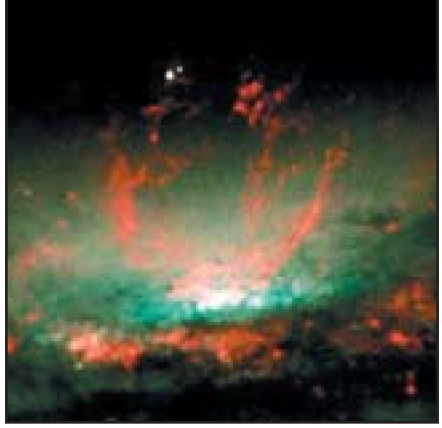
2 As the particles move around, the gravitational field changes, which causes particles to gain or lose energy.



3 Gradually the system settles down into virial equilibrium, in which the gravitational field no longer fluctuates.

DON DIXON

Supernova explosions could expel mass from low-mass galaxies so efficiently that hardly any stars would form.



that may account for the heterogeneity of the galactic dwarfs.

Although the standard picture of galaxy formation is remarkably successful, researchers are still far from working out all the processes involved. Moreover, they have yet to resolve some troubling inconsistencies. The simple picture of gas cooling inside dark matter halos faces an important problem known as the cooling catastrophe. Calculations of the cooling rates imply that the gas should have cooled briskly and pooled in the centers of halos, leaving intergalactic space virtually empty. Yet the space between galaxies is far from empty. Some extra input of energy must have prevented the gas from cooling down.

Some Feedback, Please

ANOTHER PROBLEM CONCERNS angular momentum. The amount of angular momentum imparted to protogalaxies in the models is comparable to the angular momentum that we actually see in spiral galaxies. So long as the gas retains its angular momentum, the CDM picture reproduces the observed sizes of spirals. Unfortunately, in the simulations the angular momentum leaks away. Much of it is transferred to the dark matter during galaxy mergers. As a result, the disks emerging from these simulations are a factor of 10 too small. Apparently the models are still missing an essential ingredient.

A third inconsistency has to do with the number of dwarf galaxies. Hierarchical theories predict a proliferation of low-mass dark matter halos and, by extension, dwarf galaxies. These are simply not seen. In the neighborhood of the Milky Way, the number of low-mass dwarfs is a factor of 10 to 100 lower than theories predict. Either these dark matter halos do not exist or they are present but have eluded detection because stars do not form within them.

Several solutions have been suggested for these problems. The proposals fall into two classes: either a fundamental change to the model, perhaps to the nature of dark matter [see "What's the Matter?" by George Musser; News and Analysis, SCIENTIFIC AMERICAN, May 2000], or a revision of our picture of how the cooling gas is transformed into stars. Because most astronomers are reluctant to abandon the CDM model, which works so well on scales larger than galaxies, they have concentrated on improving the treatment of star formation. Current models gloss over the process, which occurs on scales that are much smaller than a typical galaxy. Incorporating it in full is far beyond the capabilities of today's supercomputers.

Yet star formation can have profound effects on the structure of a galaxy [see "The Gas between the Stars," by Ronald J. Reynolds; SCIENTIFIC AMERICAN, January 2002]. Some astronomers think that the action of stars might actually solve all three problems at once. The energy released by stars can heat the gas, obviating the cooling catastrophe. Heating also slows the descent of gas toward the center of the galaxy and thereby reduces its tendency to transfer angular momentum to the dark matter—alleviating the angular momentum problem. And supernova explosions could expel mass from the galaxies back into the intergalactic medium [see "Colossal Galactic Explosions," by Sylvain Veilleux, Gerard Cecil and Jonathan Bland-Hawthorn; SCIENTIFIC AMERICAN, February 1996]. For the lowest-mass halos, whose escape velocity is small, the process could be so efficient that hardly any stars form, which would explain why we observe fewer dwarf galaxies than predicted.

Because our understanding of these processes is poor, the models still have a lot of wiggle room. It remains to be seen whether the problems really can be fixed or whether they indicate a need for a completely new framework. Our theory of galaxy formation will surely continue to evolve. The observational surveys under way, such as the Sloan Digital Sky Survey, will enormously improve the data on both nearby and distant galaxies. Further advances in cosmology will help constrain the initial conditions for galaxy formation. Already, precise observations of the cosmic microwave background radiation have pinned down the values of the large-scale cosmological parameters, freeing galactic modelers to focus on the small-scale intricacy. Soon we may unite the large, the small and the medium into a seamless picture of cosmic evolution. SA

GERALD CECIL ET AL. *University of North Carolina at Chapel Hill* AND NASA (*superbubble in NGC 3079*)

MORE TO EXPLORE

Cosmological Physics. John A. Peacock. Cambridge University Press, 1999.

The Formation of Ellipticals, Black Holes and Active Galactic Nuclei: A Theoretical Perspective. Guinevere Kauffmann, Stéphane Charlot and Martin G. Haehnelt in *Philosophical Transactions of the Royal Society of London*, Series A, Vol. 358, No. 1772, pages 2121–2132; July 15, 2000.

The Big Bang. Joseph Silk. W. H. Freeman and Company, 2001.

The Morphological Evolution of Galaxies. Roberto G. Abraham and Sidney van den Bergh in *Science*, Vol. 293, No. 5533, pages 1273–1278; August 17, 2001. Available at **arxiv.org/abs/astro-ph/0109358**

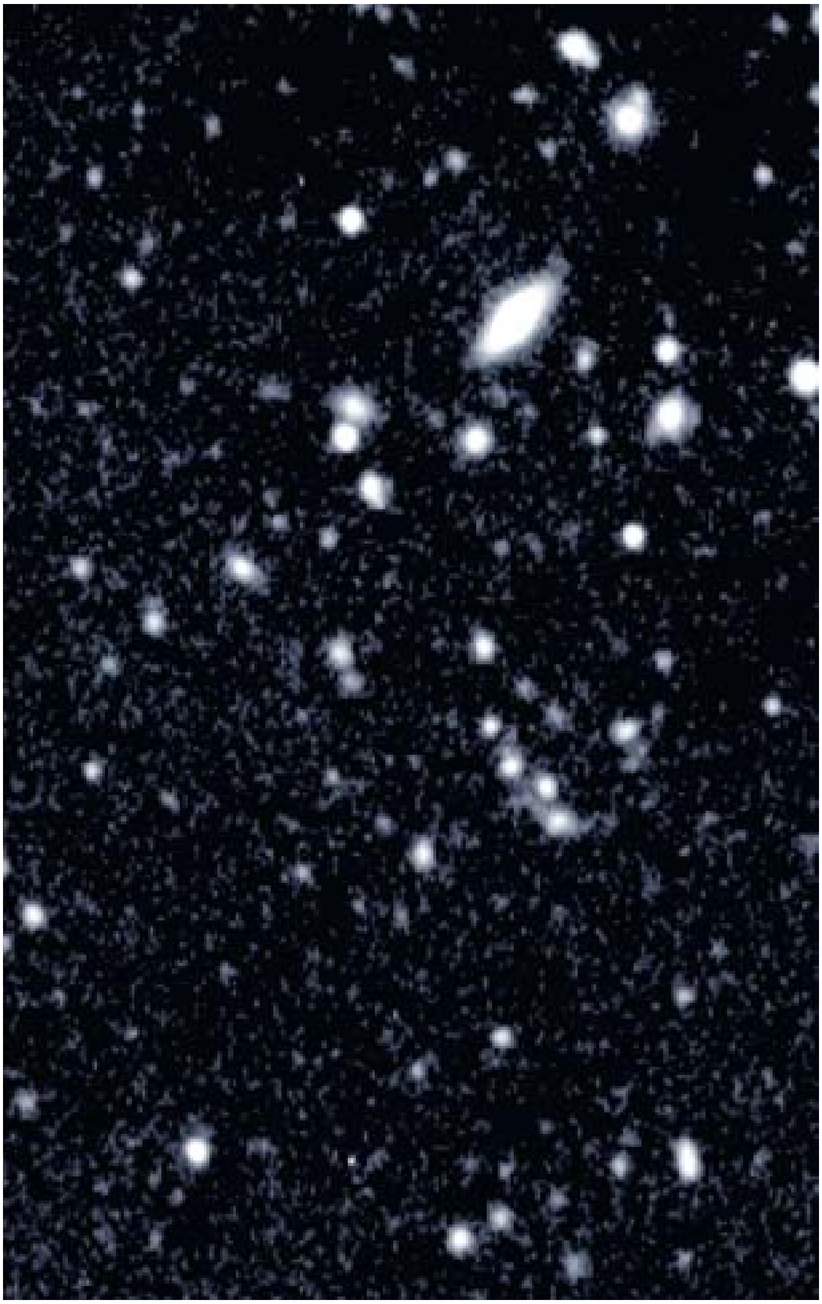
The Angular Momentum Content of Dwarf Galaxies: New Challenges for the Theory of Galaxy Formation. Frank C. van den Bosch, Andreas Burkert and Rob A. Swaters in *Monthly Notices of the Royal Astronomical Society*, Vol. 326, No. 3, pages 1205–1215; September 21, 2001. Available at **astro-ph/0105082**

New Perspectives in Astrophysical Cosmology. Martin Rees. Second edition. Cambridge University Press, 2002.

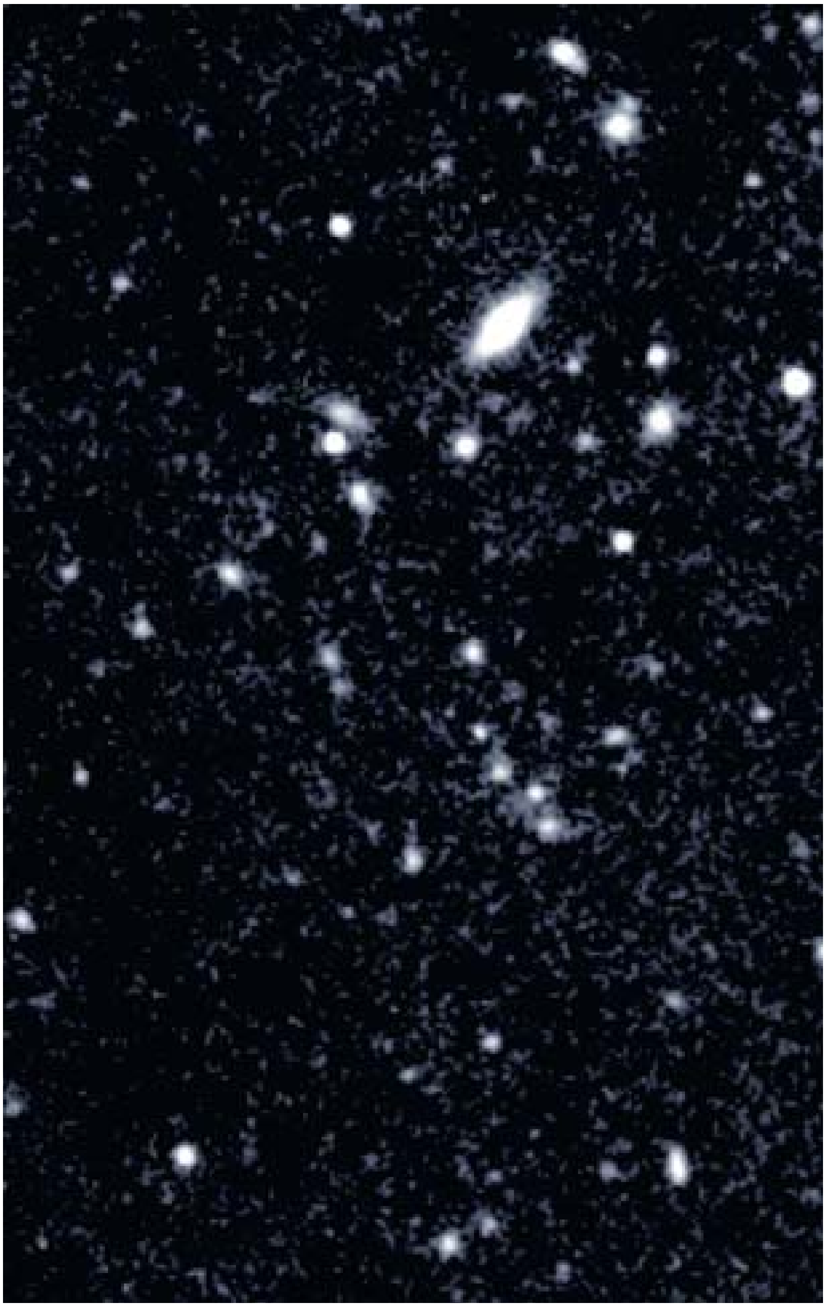
Galaxy Formation and Evolution: Recent Progress. Richard S. Ellis. Lecture given at the XIth Canary Islands Winter School of Astrophysics, "Galaxies at High Redshift" (in press). Available at **astro-ph/0102056**

EXPANSION

SURVEYING SPACETIME



WITH SUPERNOVAE



PETER CHALLIS *Harvard-Smithsonian Center for Astrophysics*

Exploding stars seen across immense distances show that the cosmic expansion may be accelerating—a sign that an exotic new form of energy could be driving the universe apart

By Craig J. Hogan, Robert P. Kirshner and Nicholas B. Suntzeff

WHERE'S THE SUPERNOVA? This pair of images, made by the authors' team using the four-meter-diameter Blanco Telescope at Cerro Tololo Inter-American Observatory in Chile, provided the first evidence of one supernova. In the image at the right, obtained three weeks after the one at the left, the supernova visibly (but subtly) alters the appearance of one of the galaxies. Can you find it? Some differences are caused by varying atmospheric conditions. To check, consult the key on page 28.

A long time ago (some five billion years),

in a galaxy far, far away (about 2,000 megaparsecs), a long-dead star exploded with a flash brighter than a billion suns. Its light spread out across space, fading and stretching with the expanding cosmos, before some of it finally reached the earth. Within 10 minutes during one dark night in 1997, a few hundred photons from this supernova landed on the mirror of a telescope in Chile. A computer at the observatory then created a digital image that showed the arrival of this tiny blip of light. Though not very impressive to look at, for us this faint spot was a thrilling sight—a new beacon for surveying space and time.

We and our colleagues around the world have tracked the arrival of light from several dozen such supernovae and used these observations to map the overall shape of the universe and to chronicle its expansion. What we and another team of astronomers have recently discerned challenges decades of conventional wisdom: it seems the universe is bigger and emptier than suspected. Moreover, its ongoing expansion is not slowing down as much as many cosmologists had anticipated; in fact, it may be speeding up.

Distant supernovae are 25 percent dimmer than was forecast, indicating an accelerating expansion of space.

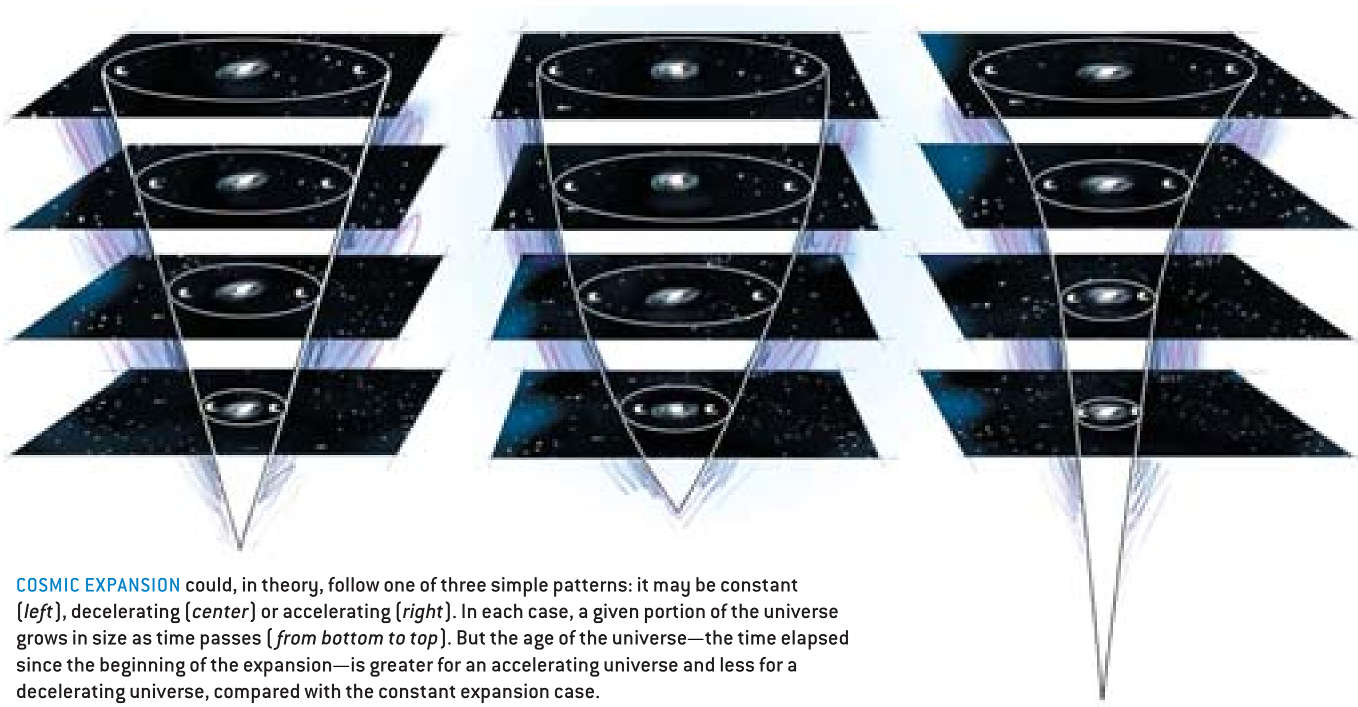
Star Warps

THE HISTORY OF COSMIC EXPANSION has been of keen interest for nearly a century, because it reflects on both the geometry of the universe and the nature of its constituents—matter, light and possibly other, more subtle forms of energy. Einstein's general theory of relativity knits together these fundamental properties of the universe and describes how they affect the motion of matter and the propagation of light, thereby offering predictions for concrete things that astronomers can actually measure.

Before the publication of Einstein's theory in 1916 and the first observations of cosmic expansion during the following decade, most scientists thought the universe stayed the same size. Indeed, Einstein himself distrusted his equations when he realized they implied a dynamic universe. But new measurements of galactic motions by Edwin P. Hubble and others left no doubt: faint, distant galaxies were flying away from the earth faster than bright, nearby ones, matching the predictions of general relativity for a universe that grows and carries galaxies farther apart. These researchers determined the outward velocities of galaxies from the shift of visible spectral lines to longer wavelengths (so-called redshifts). Though often ascribed to the Doppler effect—the phenomenon responsible for the changing pitch of a passing train whistle or car horn—the cosmological redshift is more correctly thought of as a result of the ongoing expansion of the universe, which stretches the wavelength of light passing between galaxies. Emissions from more distant objects, having traveled for a greater time, become more redshifted than radiation from nearer sources.

The technology of Hubble's day limited the initial probing of cosmic expansion to galaxies that were comparatively close. In the time it took light from these nearby galaxies to reach the earth, the universe had expanded by only a small fraction of its overall size. For such modest changes, redshift is directly proportional to distance; the fixed ratio of the two is called Hubble's constant and denotes the current rate of cosmic expansion. But astronomers have long expected that galaxies farther away would depart from this simple relation between redshift and distance, either because the pace of expansion has changed over time or because the intervening space is warped. Measuring this effect thus constitutes an important goal for cosmologists—but it is a difficult one, requiring the means to determine the distances to galaxies situated tremendously far away.

Hubble and other pioneers estimated distances to various galaxies by assuming that they all had the same intrinsic brightness. According to their logic, the ones that appeared bright were comparatively close, and the ones that appeared dim were far away. But this methodology works only crudely, because galaxies differ in their properties. And it fails entirely for distant sources—whose light takes so long to reach the earth that it reveals the faraway galaxies as they were billions of years ago (that is, in their youth)—because their intrinsic brightness could



COSMIC EXPANSION could, in theory, follow one of three simple patterns: it may be constant (*left*), decelerating (*center*) or accelerating (*right*). In each case, a given portion of the universe grows in size as time passes (*from bottom to top*). But the age of the universe—the time elapsed since the beginning of the expansion—is greater for an accelerating universe and less for a decelerating universe, compared with the constant expansion case.

have been quite different from that of more mature galaxies seen closer to home. It is difficult to disentangle these evolutionary changes from the effects of the expansion, so astronomers have long sought other "standard candles" whose intrinsic brightness is better known.

To be visible billions of light-years away, these beacons must be very bright. During the early 1970s, some cosmic surveyors tried using quasars, which are immensely energetic sources (probably powered by black holes swallowing stars and gas). But the quasars they studied proved even more diverse than galaxies and thus were of little use.

About the same time, other astronomers began exploring the idea of using supernovae—exploding stars—as standard candles for cosmological studies. That approach was controversial because supernovae, too, show wide variation in their properties. But in the past decade research by members of our team has enabled scientists to determine the intrinsic brightness of one kind of supernova—type Ia—quite precisely.

Death Star

WHAT IS A TYPE IA SUPERNOVA? Essentially, it is the blast that occurs when a dead star becomes a natural thermonuclear bomb. Spectacular as this final transformation is, the progenitor begins its life as an ordinary star, a stable ball of gas whose outer layers are held up by heat from steady nuclear reactions in its core, which convert hydrogen to helium, carbon, oxygen, neon and other elements. When the star dies, the nuclear ashes coalesce into a glowing ember, compressed by gravity to the size of the earth and a million times the density of ordinary matter.

Most such white dwarf stars simply cool and fade away, dying with a whimper. But if one is orbiting near another star, it can slurp up material from its companion and become denser and denser until a runaway thermonuclear firestorm ignites. The nuclear cataclysm blows the dwarf star apart, spewing out material at about 10,000 kilometers a second. The glow of this expanding fireball takes about three weeks to reach its maximum brightness and then declines over a period of months.

These supernovae vary slightly in their brilliance, but there is a pattern: bigger, brighter explosions last somewhat longer than fainter ones. So by monitoring how long they last, astronomers can correct for the differences and deduce their inherent brightness to within 12 percent. Over the past decade studies of nearby type Ia supernovae with modern detectors have made these flashes the best calibrated standard candles known to astronomers.

One of these candles lights up somewhere in a typical galaxy about once every 300 years. Although such stellar explosions in our own Milky Way are rare celestial events, if you monitor a few thousand other galaxies, you can expect that about one type Ia supernova will appear every month. Indeed, there are so many galaxies in the universe that, somewhere in

THE AUTHORS

CRAIG J. HOGAN, ROBERT P. KIRSHNER and *NICHOLAS B. SUNTZEFF* share a long-standing interest in big things that go bang. Hogan earned his doctorate at the University of Cambridge and is professor and vice provost for research at the University of Washington. Kirshner attained his Ph.D. at the California Institute of Technology, studying a type Ia supernova observed in 1972 (the brightest one seen since 1937). He is professor of astronomy at Harvard University and also serves as associate director of the Harvard-Smithsonian Center for Astrophysics. Suntzeff received his Ph.D. at the University of California, Santa Cruz. He works at Cerro Tololo Inter-American Observatory in La Serena, Chile, and is made of elements formed in supernovae more than five billion years ago.

DANIELS & DANIELS

the sky, supernovae bright enough to study erupt every few seconds. All astronomers have to do is find them and study them carefully. For the past few years, that effort has occupied both our research group, dubbed the "High-Z Team" (for the letter that astronomers use to denote redshift), a loose affiliation organized in 1995 by Brian P. Schmidt of Mount Stromlo and Siding Spring Observatories in Australia, and a competing collaboration called the Supernova Cosmology Project, which began in 1988 and is led by Saul Perlmutter of Lawrence Berkeley National Laboratory.

changes that might be exploding stars. Because the digital light detectors can count the number of photons in each picture element precisely, we simply subtract the first image from the second and look for significant differences from zero. Because we are checking thousands of galaxies in each image pair, we can be confident that the search of multiple pairs will find many supernovae—as long as the weather is good. Fortunately, the location of the observatory, in the foothills of the Andes on the southern fringe of Chile's Atacama Desert (one of the driest places in the world), usually provides clear skies. Betting that we will make

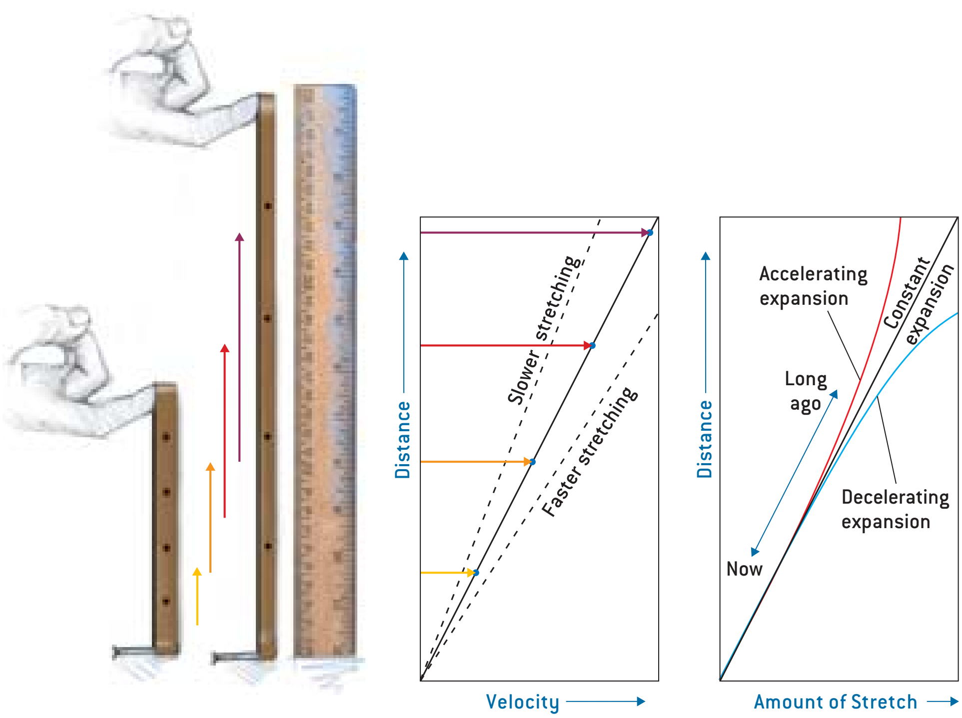
New maps of the cosmic background radiation suggest that the universe is flat and filled with dark energy.

Although the two teams have independent programs, they are exploiting the same fundamental advance: the deployment of large electronic light detectors on giant telescopes, a combination that produces digital images of faint objects over sizable swaths of the sky. A prime example of this new technology (one that has served both teams) is the Big Throughput Camera, which was developed by Gary M. Bernstein of the University of Michigan and J. Anthony Tyson of Lucent Technologies. When this camera is placed at the focus of the four-meter Blanco Telescope at Cerro Tololo Inter-American Observatory in Chile, a single exposure covers an area about as big as the full moon and creates a picture of about 5,000 galaxies in 10 minutes.

Finding distant supernovae is just a matter of taking images of the same part of the sky a few weeks apart and searching for

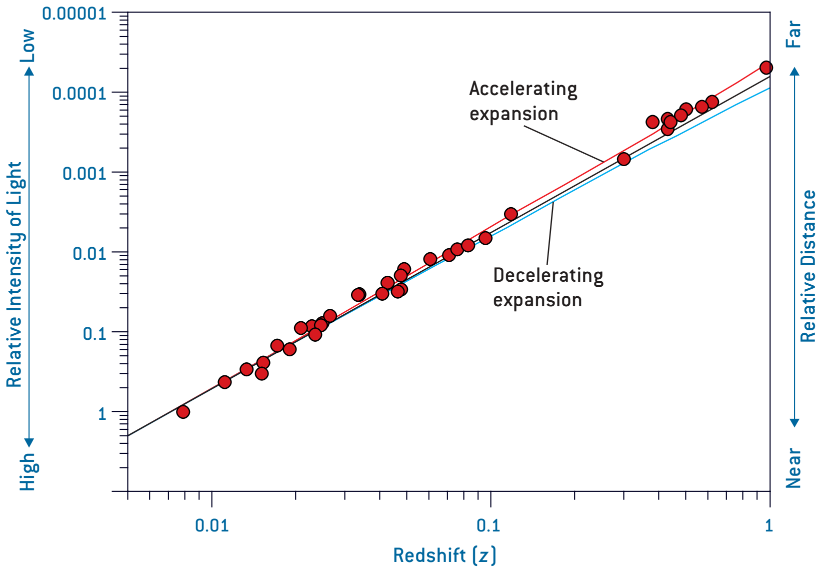
some good discoveries, we schedule observing time in advance on a battery of other telescopes around the world so that follow-up measurements can start before the supernovae fade away.

In practice, the search for exploding stars in the heavens whips up its own burst of activity on the ground, because we must acquire and compare hundreds of large, digital images at a breakneck pace. We commandeer computers scattered throughout the Cerro Tololo observatory for the tasks of aligning the images, correcting for differences in atmospheric transparency and image size, and subtracting the two scans. If all goes well, most of the galaxies disappear, leaving just a little visual "noise" in the difference of the two images. Larger signals indicate some new or changing object, such as variable stars, quasars, asteroids—and in a few cases, supernovae.



RUBBER-BAND EXPERIMENT shows the linear relation between recession velocity and distance. Here two snapshots are shown of a rubber band pulled upward at a certain rate. The velocity of different points marked on the band is given by the length of the colored arrows. For example, the point closest to the origin moves the least during the interval between snapshots, so its velocity is the smallest (*yellow arrow*). In contrast, the farthest point moves the most, so its velocity is the highest (*violet arrow*). The slope of the resulting line is the rate of expansion (*left graph*). If the rate changes over time, the slope, too, will change (*right graph*). Earlier times plot toward the upper right, because light from more distant objects takes longer to reach the earth, the origin of the plot. If the rate was slower in the past—indicating that the expansion has been accelerating—the line will curve upward (*red line*). If the rate was faster—as in a decelerating expansion—it will curve downward (*blue line*).

DANIELS & DANIELS (*illustration*); DAVID SCHNEIDER (*graphs*)



SUPERNOVA OBSERVATIONS by the authors' team (*red dots*) deviate slightly but significantly from the pattern that many theoreticians expected—namely, a fairly rapid deceleration (*blue line*) that should occur if the universe is "flat" and has no cosmological constant. These observations indicate that the universe has only 20 percent of the matter necessary to make it flat, because it is decelerating more slowly than predicted (*black line*). The measurements even suggest that expansion is accelerating, perhaps because of a nonzero cosmological constant (*red line*).

DAVID SCHNEIDER; SOURCE: CRAIG J. HOGAN, ROBERT P. KIRSHNER AND NICHOLAS B. SUNTZEFF

Our software records the position of new objects and attempts to identify which are truly supernovae. But the automated tests are imperfect, and we must scrutinize the images by eye to decide whether a putative supernova is real. Because we must immediately pursue our discoveries with other telescopes, the analysis must be done quickly. During these exhausting times, the observatory becomes a sweatshop of astronomers and visiting students, who work around the clock for days at a stretch, sustained by enthusiasm and Chilean pizza.

We next target the best supernova candidates with the largest optical instruments in the world, the Keck telescopes in Hawaii. These critical observations establish whether the objects discovered are in fact type Ia supernovae, gauge their intrinsic brightness more exactly and determine their redshifts.

On the Dark Side

OTHERS IN OUR GROUP, working with telescopes in Australia, Chile and the U.S., also follow these supernovae to track how their brilliance peaks and then slowly fades. The observing campaign for a single supernova spans months, and the final analysis often has to wait a year or more, when the light of the exploded star has all but disappeared, so we can obtain a good image of its host galaxy. We use this final view to subtract the constant glow of the galaxy from the images of the supernova. Our best measurements come from the Hubble Space Telescope, which captures such fine details that the exploding star stands out distinctly from its host galaxy.

The two teams have now studied a total of a few score high-redshift supernovae, ones that erupted between four billion and seven billion years ago, when the universe was between one half and two thirds of its present age. Both groups were hit with a major surprise: the supernovae are fainter than expected. The difference is slight, the distant supernovae being, on average, only 25 percent dimmer than forecast. But this result is enough to call long-standing cosmological theories into question.

Before drawing any sweeping conclusions, astronomers on both teams have been asking themselves whether there is a prosaic explanation for the relative dimness of these distant supernovae. One culprit could be murkiness caused by cosmic dust, which might screen out some of the light. We think we can discount this possibility, however, because dust grains would tend to filter out blue light more than red, causing the supernovae to appear redder than they really are (in the same way that atmospheric dust colors the setting sun). We observe no such alteration. Also, we would expect that cosmic dust, unless it is spread very smoothly throughout space, would introduce a large amount of variation in the measurements, which we do not see either.

Another possible disturbance is gravitational lensing, the bending of light rays as they skirt galaxies en route. Such lensing occasionally causes brightening, but most often it causes demagnification and thus can contribute to the dimness of distant supernovae. Yet calculations show that this effect becomes important only for sources located even farther away than the supernovae we are studying, so we can dismiss this complication as well.

Finally, we worried that the distant supernovae are somehow different from the nearby ones, perhaps forming from younger stars that contain fewer heavy elements than is typical in more mature galaxies. Although we cannot rule out this possibility, our analysis already tries to take such differences into account. These adjustments appear to work well when we apply them to nearby galaxies, which range widely in age, makeup and the kinds of supernovae seen.

Because none of these mundane effects fits the new observations, we and many other scientists are now led to think that the unexpected faintness of distant supernovae is indeed caused by the structure of the cosmos. Two different properties of space and of time might be contributing.

First, space might have negative curvature. Such warping is easier to comprehend with a two-dimensional analogy. Creatures living in a perfectly flat, two-dimensional world (like the characters in Edwin A. Abbott's classic novel *Flatland*) would find that a circle of radius r has a circumference of exactly $2\pi r$. But if their world were subtly bent into a saddle shape, it would have a slight negative curvature. The two-dimensional residents of Saddleland might be oblivious to this curvature until they measured a large circle of some set radius and discovered that its circumference was greater than $2\pi r$.

If the **cosmos expanded** more slowly in the past, as supernovae indicate, the age of the universe must be **revised upward**.

Most cosmologists have assumed, for various theoretical reasons, that our three-dimensional space, like Flatland, is not curved. But if it had negative curvature, the large sphere of radiation given off by an ancient supernova would have a greater area than it does in geometrically flat space, making the source appear strangely faint.

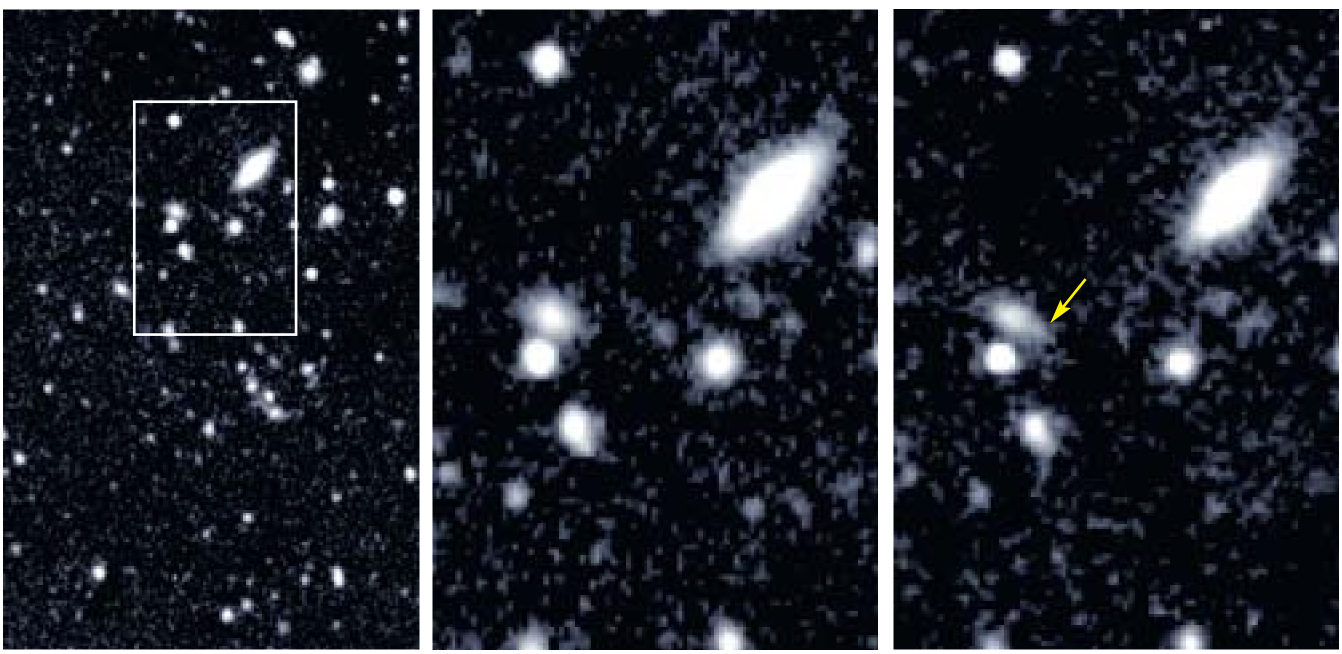
A second explanation for the unexpected dimness of distant supernovae is that they are farther away than their redshifts suggest. Viewed another way, supernovae located at these enormous distances seem to have less redshift than anticipated. To account for the smaller redshift, cosmologists postulate that the universe must have expanded more slowly in the past than they had expected, giving less of an overall stretch to the universe and to the light traveling within it.

The Force

WHAT IS THE SIGNIFICANCE of the cosmic expansion slowing less quickly than previously thought? If the universe is made of normal matter, gravity must steadily slow the expansion. Little slowing, as indicated by the supernovae measurements, thus implies that the overall density of matter in the universe is low.

Although this conclusion undermines theoretical preconceptions, it agrees with several lines of evidence. For example, astronomers have noted that certain stars appear to be older than the accepted age of the universe—a clear impossibility. But if the cosmos expanded more slowly in the past, as the supernovae now indicate, the age of the universe must be revised upward, which may resolve the conundrum. The new results also accord with other recent attempts to ascertain the total amount of matter, such as studies of galaxy clusters [see "The Evolution of Galaxy Clusters," by J. Patrick Henry, Ulrich G. Briel and Hans Böhringer; SCIENTIFIC AMERICAN, December 1998].

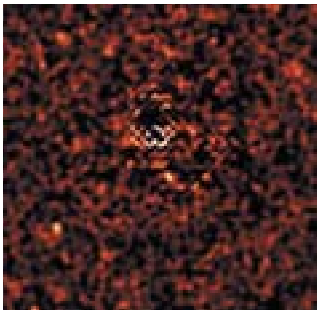
What does the new understanding of the density of matter in the universe say about its curvature? According to the principles of general relativity, curvature and deceleration are connected. To paraphrase John A. Wheeler, formerly at Princeton University: matter tells spacetime how to curve, and spacetime tells matter how to move. A small density of matter implies negative curvature as well as little slowing. If the universe is nearly empty, these two dimming effects are both near their theoretical maximum.

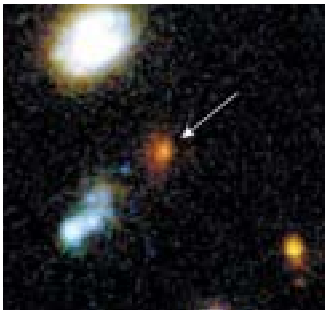


DISTANT SUPERNOVA, with a redshift of $z = 0.66$, appears by the arrow. The explosion of this star affects just a few picture elements in the image taken after the event.

PETER CHALLIS *Harvard-Smithsonian Center for Astrophysics*

IN APRIL 2001 NASA's Hubble Space Telescope captured the most distant supernova ever seen (*above* and *below, right*), which erupted 10 billion years ago. By examining its glow (*below, left*), cosmologists have found more convincing evidence that repulsive, dark energy is accelerating the universe's expansion.





NASA AND A. RIESS *Space Telescope Science Institute*

The big surprise is that the supernovae we see are fainter than predicted even for a nearly empty universe (which has maximum negative curvature). Taken at face value, our observations appear to require that expansion is actually accelerating with time. A universe composed only of normal matter cannot grow in this fashion, because its gravity is always attractive. Yet according to Einstein's theory, the expansion can speed up if an exotic form of energy fills empty space everywhere. This strange "dark energy" is embodied in Einstein's equations as the so-called cosmological constant. Unlike ordinary forms of mass and energy, the dark energy adds gravity that is repulsive and can drive the universe apart at ever increasing speeds [see "Cosmological Antigravity," on page 30]. Once we admit this extraordinary possibility, we can explain our observations perfectly, even assuming the flat geometry beloved by theorists.

Indeed, studies of a completely different kind—sky maps of the cosmic background radiation—have recently uncovered new and compelling evidence for a flat average geometry. Sound waves in the radiation matter plasma of the early universe, whose physical size can be computed from first principles, produce a blotchy pattern of anisotropy on the sky. The observed angular size of the pattern shows that the geometry is flat to high precision, implying that the radius of the whole cosmic hypersphere is very much larger than the piece of the universe we can observe (much like a small piece of the curved the earth seems flat). This flat geometry requires a total density of mass energy much greater than the total estimated density of normally gravitating matter, providing independent, if indirect, evidence that most of the stuff of the universe is made of exotic dark energy. It is striking that a multitude of increasingly precise independent techniques converge on a concordant cosmology with this deeply surprising new ingredient.

Evidence for a strange form of energy imparting a repulsive gravitational force is the most interesting result we could have hoped for, yet it is so astonishing that we and others remain suitably skeptical. Fortunately, advances in the technology available to astronomers, such as new infrared detectors and the Next Generation Space Telescope, will soon permit us to test our conclusions by offering greater precision and reliability. These marvelous instruments will also allow us to perceive even fainter beacons that flared still longer ago in galaxies that are much, much farther away. SA

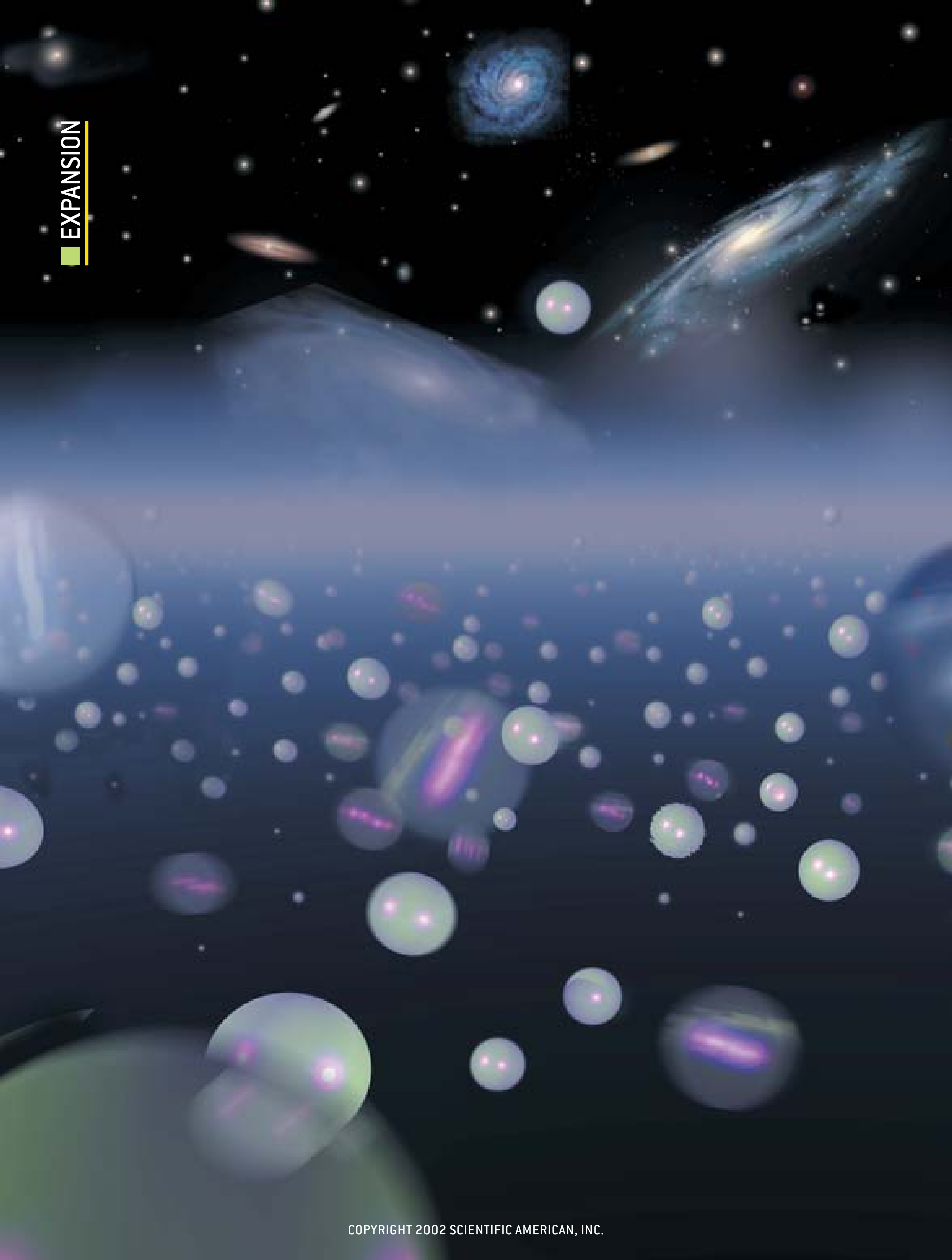
MORE TO EXPLORE

The Little Book of the Big Bang. Craig J. Hogan. Springer-Verlag, 1998.

Discovery of a Supernova Explosion at Half the Age of the Universe. S. Perlmutter, G. Aldering, M. Della Valle, S. Deustua, R. S. Ellis, S. Fabbro, A. Fruchter, G. Goldhaber, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, R. A. Knop, C. Lidman, R. G. McMahon, Peter Nugent, R. Pain, N. Panagia, C. R. Pennypacker, P. Ruiz-Lapuente, B. Schaefer and N. Walton (The Supernova Cosmology Project) in *Nature,* Vol. 391, pages 51–54; January 1, 1998. Preprint available at **xxx.lanl.gov/abs/astro-ph/9712212**

Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. Adam G. Riess, Alexei V. Filippenko, Peter Challis, Alejandro Clocchiattia, Alan Diercks, Peter M. Garnavich, Ron L. Gilliland, Craig J. Hogan, Saurabh Jha, Robert P. Kirshner, B. Leibundgut, M. M. Phillips, David Reiss, Brian P. Schmidt, Robert A. Schommer, R. Chris Smith, J. Spyromilio, Christopher Stubbs, Nicholas B. Suntzeff and John Tonry in *Astronomical Journal,* Vol. 116, No. 3, pages 1009–1038; September 1998. Preprint at **xxx.lanl.gov/abs/astro-ph/9805201**

Additional information on supernova searches is available at **cfa-www.harvard.edu/cfa/oir/Research/supernova/HighZ.html** and **www.supernova.lbl.gov/**



THE LONG-DERIDED COSMOLOGICAL CONSTANT— A CONTRIVANCE OF ALBERT EINSTEIN'S— MAY EXPLAIN CHANGES IN THE EXPANSION RATE OF THE UNIVERSE

Cosmological ANTIGRAVITY

BY LAWRENCE M. KRAUSS

SO-CALLED EMPTY SPACE is actually filled with elementary particles that pop in and out of existence too quickly to be detected directly. Their presence is the consequence of a basic principle of quantum mechanics combined with special relativity: nothing is exact, not even nothingness. The aggregate energy represented by these virtual particles, like other forms of energy, could exert a gravitational force, which could be either attractive or repulsive depending on physical principles that are not yet understood. On macroscopic scales the energy could act as the cosmological constant proposed by Albert Einstein.

Novelist and social critic George Orwell wrote in 1946, "To see what is in front of one's nose requires a constant struggle." These words aptly describe the workings of modern cosmology. The universe is all around us—we are part of it—yet scientists must sometimes look halfway across it to understand the processes that led to our existence on the earth. And although researchers believe that the underlying principles of nature are simple, unraveling them is another matter. The clues in the sky can be subtle. Orwell's adage is doubly true for cosmologists grappling with the recent observations of exploding stars hundreds of millions of light-years away. Contrary to most expectations, they are finding that the expansion of the universe may not be slowing down but rather speeding up.

Astronomers have known that the visible universe is expanding since at least 1929, when Edwin P. Hubble demonstrated that distant galaxies are moving apart as they would if the entire cosmos were uniformly swelling in size. These outward motions are counteracted by the collective gravity of galaxy clusters and all the planets, stars, gas and dust they contain. Even the minuscule gravitational pull of, say, a paper clip retards cosmic expansion by a slight amount. A decade ago a congruence of theory and observations suggested that there were enough paper clips and other matter in the universe to almost, but never quite, halt the expansion. In the geometric terms that Albert Einstein encouraged cosmologists to adopt, the universe seemed to be "flat."

The flat universe is an intermediate between two other plausible geometries, called "open" and "closed." In a cosmos where matter does battle with the outward impulse from the big bang, the open case represents the victory of expansion: the universe would go on expanding forever. In the closed case, gravity would have the upper hand, and the universe would eventually collapse again, ending in a fiery "big crunch." The open, closed and flat scenarios are analogous to launching a rocket faster than, slower than or exactly at the earth's escape velocity—the speed necessary to overcome the planet's gravitational attraction.

That we live in a flat universe, the perfect balance of power, is one of the hallmark predictions of standard inflationary theory, which postulates a very early period of rapid expansion to reconcile several paradoxes in the conventional formulation of the big bang. Although the visible contents of the cosmos are clearly not enough to make the universe flat, celestial dynamics indicate that there is far more matter than meets the eye. Most of the material in galaxies and assemblages of galaxies must be invisible to telescopes. Over a decade ago I applied the term "quintessence" to this so-called dark matter, borrowing a term Aristotle used for the ether—the invisible material supposed to permeate all of space [see "Dark Matter in the Universe," by Lawrence M. Krauss; SCIENTIFIC AMERICAN, December 1986].

Yet an overwhelming body of evidence now implies that even the unseen matter is not enough to produce a flat universe. If that is so, its main constituents cannot be visible matter, dark matter or radiation. Instead the universe must be composed largely of an even more ethereal form of energy that inhabits empty space, including that which is in front of our noses.

Fatal Attraction

The idea of such energy has a long and checkered history, which began when Einstein completed his general theory of relativity, more than a decade before Hubble's convincing demonstration that the universe is expanding. By tying together space, time and matter, relativity promised what had previously been impossible: a scientific understanding not merely of the dynamics of objects within the universe but of the universe itself. There was only one problem. Unlike other fundamental forces felt by matter, gravity is universally attractive—it only pulls; it cannot push. The unrelenting gravitational attraction of matter could cause the universe to collapse eventually. So Einstein, who presumed the universe to be static and stable, added an extra term to his equations, a "cosmological term," which could stabilize the universe by producing a new long-range force throughout space. If its value were positive, the term would represent a repulsive force—a kind of antigravity that could hold the universe up under its own weight.

Alas, within five years Einstein abandoned this kludge, which he associated with his "biggest blunder." The stability offered by the term turned out to be illusory, and, more important, evidence had begun to mount that the universe is expanding. As early as 1923, Einstein wrote in a letter to mathematician Hermann Weyl that "if there is no quasi-static world, then away with the cosmological term!" Like the ether before it, the term appeared to be headed for the dustbin of history.

Physicists were happy to do without such an intrusion. In

THE AUTHOR

LAWRENCE M. KRAUSS works at the interface of physics and astronomy. He studies the workings of stars, black holes, gravitational lenses and the early universe in order to shed light on particle physics beyond the current Standard Model, including the unification of forces, quantum gravity and explanations for dark matter. Krauss is currently chair of the physics department at Case Western Reserve University. He is the author of six books, most recently *Atom: A Single Oxygen Atom's Odyssey from the Big Bang to Life on Earth ... and Beyond* (Back Bay Books, 2002).

ALFRED T. KAMAJIAN (*preceding pages*)

Types of Matter

CONTENTS OF THE UNIVERSE include billions of galaxies, each one containing an equally mind-boggling number of stars. Yet the bulk of matter seems to consist of dark matter, whose identity is still uncertain. The cosmological constant, if its existence is confirmed, would act like a yet more exotic form of dark energy on cosmological scales. The quantity omega (Ω) is the ratio of the density of matter or energy to the density required for flatness. *—L.M.K.*

TYPE	LIKELY COMPOSITION	MAIN EVIDENCE	APPROXIMATE CONTRIBUTION TO Ω
VISIBLE MATTER	Ordinary matter (composed mainly of protons and neutrons) that forms stars, dust and gas	Telescope observations	0.01
BARYONIC DARK MATTER	Ordinary matter that is too dim to see, perhaps brown or black dwarfs (massive compact halo objects, or MACHOs)	Big bang nucleosynthesis calculations and observed deuterium abundance	0.05
NONBARYONIC DARK MATTER	Exotic particles such as "axions," neutrinos with mass or weakly interacting massive particles (WIMPs)	Gravity of visible matter is insufficient to account for orbital speeds of stars within galaxies and of galaxies within clusters	0.3
COSMOLOGICAL DARK ENERGY	Cosmological constant (energy of empty space)	Microwave background suggests cosmos is flat, but there is not enough baryonic or nonbaryonic matter to make it so	0.7

the general theory of relativity, the source of gravitational forces (whether attractive or repulsive) is energy. Matter is simply one form of energy. But Einstein's cosmological term is distinct. The energy associated with it does not depend on position or time—hence the name "cosmological constant." The force caused by the constant operates even in the complete absence of matter or radiation. Therefore, its source must be a curious energy that resides in empty space. The cosmological constant, like the ether, endows the void with an almost metaphysical aura. With its demise, nature was once again reasonable.

Or was it? In the 1930s glimmers of the cosmological constant arose in a completely independent context: the effort to combine the laws of quantum mechanics with Einstein's special theory of relativity. Physicists Paul A. M. Dirac and later Richard Feynman, Julian S. Schwinger and Shinichiro Tomonaga showed that empty space was more complicated than anyone had previously imagined. Elementary particles, it turned out, can spontaneously pop out of nothingness and disappear again, if they do so for a time so short that one cannot measure them directly [see "Exploiting Zero-Point Energy," by Philip Yam; SCIENTIFIC AMERICAN, December 1997]. Such virtual particles, as they are called, may appear as far-fetched as angels sitting on the head of a pin. But there is a difference. The unseen particles produce measurable effects, such as alterations to the energy levels of atoms as well as forces between nearby metal plates. The theory of virtual particles agrees with observations to nine decimal places. (Angels, in contrast, normally have no discernible effect on either atoms or plates.) Like it or not, empty space is not empty after all.

Virtual Reality

IF VIRTUAL PARTICLES can change the properties of atoms, might they also affect the expansion of the universe? In 1967 Russian astrophysicist Yakov B. Zeldovich showed that the energy of virtual particles should act precisely as the energy associated with a cosmological constant. But there was a serious problem. Quantum theory predicts a whole spectrum of virtual particles, spanning every possible wavelength. When physicists add up all the effects, the total energy comes out infinite. Even if theorists ignore quantum effects smaller than a certain wavelength—for which poorly understood quantum gravitational effects presumably alter things—the calculated vacuum energy is roughly 120 orders of magnitude larger than the energy contained in all the matter in the universe.

What would be the effect of such a humongous cosmological constant? Taking a cue from Orwell's maxim, you can easily put an observational limit on its value. Hold out your hand and look at your fingers. If the constant were as large as quantum theory naively suggests, the space between your eyes and your hand would expand so rapidly that the light from your hand would never reach your eyes. To see what is in front of your face would be a constant struggle (so to speak), and you would always lose. The fact that you can see anything at all means that the energy of empty space cannot be large. And the fact that we can see not only to the ends of our arms but also to the far reaches of the universe puts an even more stringent limit on the cosmological constant: almost 120 orders of magnitude smaller than the estimate mentioned above. The discrepancy between theory and observation is the most perplexing quantitative puzzle in physics [see "The Mystery of the Cosmological Constant," by Larry Abbott; SCIENTIFIC AMERICAN, May 1988].

The simplest conclusion is that some as yet undiscovered physical law causes the cosmological constant to vanish. But as much as theorists might like the constant to go away, various astronomical observations—of the age of the universe, the density of matter and the nature of cosmic structures—all independently suggest that it may be here to stay.

Determining the age of the universe is one of the long-standing issues of modern cosmology. By measuring the velocities of galaxies, astronomers can calculate how long it took them to arrive at their present positions, assuming they all started out at the same place. For a first approximation, one can ignore the deceleration caused by gravity. Then the universe would expand at a constant speed and the time interval would just be the ratio of the distance between galaxies to their measured speed of separation—that is, the reciprocal of the famous Hubble constant. The higher the value of the Hubble constant, the faster the expansion rate and hence the younger the universe.

Hubble's first estimate of his eponymous constant was almost 500 kilometers per second per megaparsec—which would mean that two galaxies separated by a distance of one megaparsec (about three million light-years) are moving apart, on average, at 500 kilometers per second. This value would imply a cosmic age of about two billion years, which is in painful contradiction with the known age of the earth—about four billion years. When the gravitational attraction of matter is taken into account, the analysis predicts that objects moved faster early on, taking even less time to get to their present positions than if their speed had been constant. This refinement reduces the age estimate by one third, unfortunately worsening the discrepancy.

Over the past seven decades, astronomers have improved their determination of the expansion rate, but the tension between the calculated age of the universe and the age of objects within it has persisted. In the past decade, with the launch of the Hubble Space Telescope and the development of new observational techniques, disparate measurements of the Hubble constant are finally beginning to converge. Wendy L. Freedman of the Carnegie Observatories and her colleagues have inferred a value of 73 kilometers per second per megaparsec (with a most likely range, depending on experimental error, of 65 to 81). These results put the upper limit on the age of a flat universe at about 10 billion years.

The cosmological constant provides **much of the force** that holds the universe up **under its own weight.**

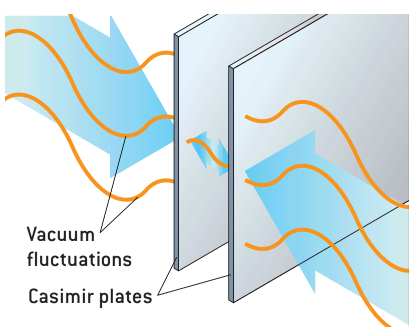
The Age Crisis

IS THAT VALUE OLD ENOUGH? It depends on the age of the oldest objects that astronomers can date. Among the most ancient stars in our galaxy are those found in tight groups known as globular clusters, some of which are located in the outskirts of our galaxy and are thus thought to have formed before the rest of the Milky Way. Estimates of their age, based on calculations of how fast stars burn their nuclear fuel, traditionally ranged from 15 billion to 20 billion years. Such objects appeared to be older than the universe.

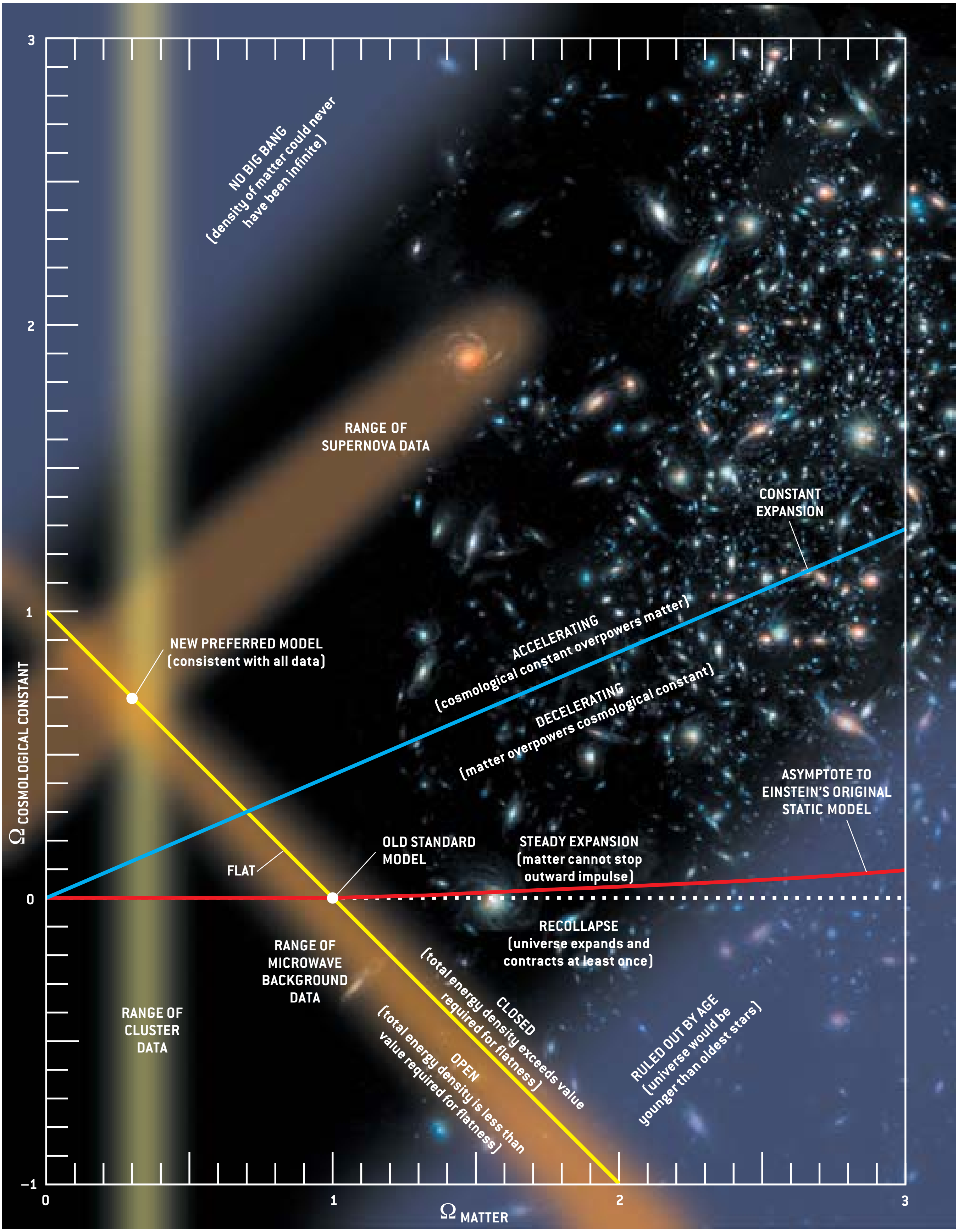
To determine whether this age conflict was the fault of cosmology or of stellar modeling, in 1995 my colleagues—Brian C. Chaboyer, then at the Canadian Institute of Theoretical Astrophysics, Pierre Demarque of Yale University and Peter J. Kernan of Case Western Reserve University—and I reassessed the globular cluster ages. We simulated the life cycles of three million different stars whose properties spanned the existing un-

PARTICLES IN EMPTY SPACE

DEMONSTRATION OF CASIMIR EFFECT is one way that physicists have corroborated the theory that space is filled with fleeting virtual particles. The Casimir effect generates forces between metal objects—for instance, an attractive force between parallel metal plates (*near right*). Loosely speaking, the finite spacing of the plates prevents virtual particles larger than a certain wavelength from materializing in the gap. Therefore, there are more particles outside the plates than between them, an imbalance that pushes the plates together (*far right*). The Casimir effect has a distinctive dependence on the shape of the plates, which allows physicists to tease it out from other forces of nature. —*L.M.K.*



JARED SCHNEIDMAN DESIGN (*illustration*); STEVE K. LAMOREAUX *Los Alamos National Laboratory* (*photograph*)



DON DIXON (*illustration*); DMITRY KRASNY (*chart*); SOURCE: LAWRENCE M. KRAUSS AND MAX TEGMARK

MAP OF MODELS shows how the unfolding of the universe depends on two key cosmological quantities: the average density of matter (*horizontal axis*) and the density of energy in the cosmological constant (*vertical axis*). Their values, given here in standard cosmological units, have three distinct effects. First, their sum (which represents the total cosmic energy content) determines the geometry of spacetime (*yellow line*). Second, their difference (which represents the relative strength of expansion and gravity) determines how the expansion rate changes over time (*blue line*). These two effects have been probed by recent observations (*shaded regions*). The third, a balance of the two densities, determines the fate of the universe (*red line*). The three effects have many permutations—unlike the view in which the cosmological constant is assumed to be zero and there are only two possible outcomes.

certainties and then compared our model stars with those in globular clusters. The oldest, we concluded, could be as young as 12.5 billion years old, which was still at odds with the age of a flat, matter-dominated universe.

But two years ago the Hipparcos satellite, launched by the European Space Agency to measure the locations of more than 100,000 nearby stars, revised the distances to these stars and, indirectly, to globular clusters. The new distances affected estimates of their brightness and forced us to redo our analysis, because brightness determines the rate at which stars consume fuel and hence their life spans. Although the newly derived ages were somewhat smaller, the most recent analysis by our group puts a best-fit age of the universe at 13.4 billion years, with a lower limit of 11.2 billion years, clearly in conflict with the upper limit for a flat, matter-dominated universe.

A lower density of matter, signifying an open universe with slower deceleration, would ease the tension somewhat. Even so, the only way to lift the age above 12.5 billion years would be to consider a universe dominated not by matter but by a cosmological constant. The resulting repulsive force would cause the Hubble expansion to accelerate over time. Galaxies would have been moving apart slower than they are today, taking longer to reach their present separation, so the universe would be older.

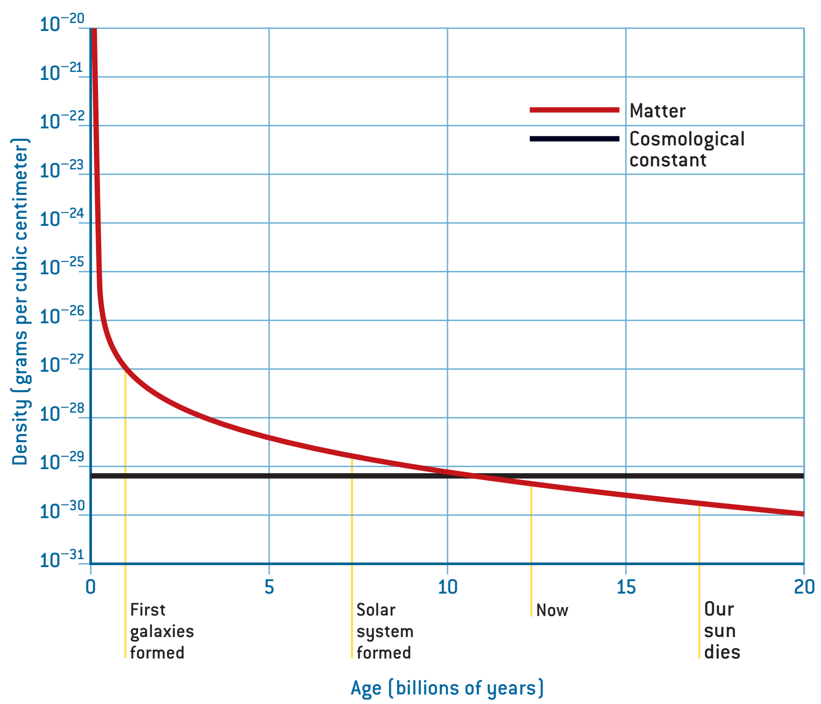
Meanwhile other pillars of observational cosmology have recently been shaken, too. As astronomers have used the latest technology to survey ever larger regions of the cosmos, their ability to tally up its contents has improved. Now the case is compelling that the total amount of matter is insufficient to yield a flat universe.

This cosmic census first involves calculations of the synthesis of elements by the big bang. The light elements in the cosmos—hydrogen and helium and their rarer isotopes, such as deuterium—were created in the early universe in relative amounts that depended on the number of available protons and neutrons, the constituents of normal matter. Thus, by comparing the abundances of the various isotopes, astronomers can deduce the total amount of ordinary matter that was produced in the big bang. (There could, of course, also be other matter not composed of protons and neutrons.)

The relevant observations took a big step forward in 1996, when David R. Tytler and Scott Burles of the University of California at San Diego and their colleagues measured the primordial abundance of deuterium using absorption of quasar light by intergalactic hydrogen clouds. Because these clouds have never contained stars, their deuterium could have been created only by the big bang. Tytler and Burles's finding implies that the average density of ordinary matter is between 4 and 7 percent of the amount needed for the universe to be flat.

Astronomers have also probed the density of matter by studying the largest gravitationally bound objects in the universe: clusters of galaxies. These groupings of hundreds of galaxies account for almost all visible matter. Most of their luminous content takes the form of hot intergalactic gas, which emits x-rays. The temperature of this gas, inferred from the spectrum of the x-rays, depends on the total mass of the cluster: in more massive clusters, the gravity is stronger and hence the pressure that supports the gas against gravity must be larger, which drives the temperature higher. In 1993 Simon D. M. White, now at the Max Planck Institute for Astrophysics in Garching, Germany, and his colleagues compiled information about several different clusters to argue that luminous matter accounted for between 10 and 20 percent of the total mass of the objects. When combined with the measurements of deuterium, these results imply that the total density of clustered matter—including protons and neutrons as well as more exotic particles such as certain dark matter candidates—is at most 60 percent of that required to flatten the universe.

AVERAGE DENSITY OF THE UNIVERSE



COSMIC COINCIDENCE is one of many mysteries swirling about the cosmological constant. The average density of ordinary matter decreases as the universe expands (*red*). The equivalent density represented by the cosmological constant is fixed (*black*). So why, despite these opposite behaviors, do the two have nearly the same value today? The consonance is either happenstance, a precondition for human existence (an appeal to the weak anthropic principle) or an indication of a mechanism not currently envisaged. *—L.M.K.*

GEORGE MUSSER AND DMITRY KRASNY

The Fate of the Universe

THE COSMOLOGICAL CONSTANT changes the usual simple picture of the future of the universe. Traditionally, cosmology has predicted two possible outcomes that depend on the geometry of the universe or, equivalently, on the average density of matter. If the density of a matter-filled universe exceeds a certain critical value, it is "closed," in which case it will eventually stop expanding, start contracting and ultimately vanish in a fiery apocalypse. If the density is less than the critical value, the universe is "open" and will expand forever. A "flat" universe, for which the density equals the critical value, also will expand forever but at an ever slower rate.

Yet these scenarios assume that the cosmological constant equals zero. If not, it—rather than matter—may control the ultimate fate of the universe. The reason is that the constant, by definition, represents a fixed density of energy in space. Matter cannot compete: a doubling in radius dilutes its density eightfold. In an expanding universe the energy density associated with a cosmological constant must win out. If the constant has a positive value, it generates a long-range repulsive force in space, and the universe will continue to expand even if the total energy density in matter and in space exceeds the critical value. (Large negative values of the constant are ruled out because the resulting attractive force would already have brought the universe to an end.)

Even this new prediction for eternal expansion assumes that the constant is indeed constant, as general relativity suggests that it should be. If in fact the energy density of empty space does vary with time, the fate of the universe will depend on how it does so. And there may be a precedent for such changes—namely, the inflationary expansion in the primordial universe. Perhaps the universe is just now entering a new era of inflation, one that may eventually come to an end. *—L.M.K.*

A third set of observations, ones that also bear on the distribution of matter at the largest scales, supports the view that the universe has too little mass to make it flat. Perhaps no other subfield of cosmology has advanced so much in the past 20 years as the understanding of the origin and nature of cosmic structures. Astronomers had long assumed that galaxies coalesced from slight concentrations of matter in the early universe, but no one knew what would have produced such undulations. The development of the inflationary theory in the 1980s provided the first plausible mechanism—namely, the enlargement of quantum fluctuations to macroscopic size.

Numerical simulations of the growth of structures following inflation have shown that if dark matter was not made from protons and neutrons but from some other type of particle (such as so-called WIMPs), tiny ripples in the cosmic microwave background radiation could grow into the structures now seen. Moreover, concentrations of matter should still be evolving into clusters of galaxies if the overall density of matter is high. The relatively slow growth of the number of rich clusters over the recent history of the universe suggests that the density of matter is less than 50 percent of that required for a flat universe [see "The Evolution of Galaxy Clusters," by J. Patrick Henry, Ulrich G. Briel and Hans Böhringer; SCIENTIFIC AMERICAN, December 1998].

Nothing Matters

THESE MANY FINDINGS that the universe has too little matter to make it flat have become convincing enough to overcome the strong theoretical prejudice against this possibility. Two interpretations are possible: either the universe is open, or it is made flat by some additional form of energy that is not associated with ordinary matter. To distinguish between these alternatives, astronomers have been pushing to measure the microwave background at high resolution. The cosmic microwave background—the literal afterglow of the big bang—emanates from a "last scattering" surface located more than 12 billion light-years away from us. The surface represents a time when the universe first cooled sufficiently so that the previously ionized plasma of protons and electrons could combine to form neutral hydrogen, which is transparent to radiation.

When we measure the CMB radiation today in different directions, we are measuring regions that were far enough apart that they could not have been in casual contact back then. Regions separated by less than about a degree could have been traversed at the speed of light over the 300,000 years or so it took the radiation gas to cool. This angular scale should leave a remnant imprint in the CMB "map" that current detectors measure. The actual angular scale associated with this distance, however, depends on the geometry of the universe. In a flat universe, light rays travel in straight lines as one traces them back to their source. In an open universe, light rays diverge as one traces them back, and in a closed universe they converge. In an open universe, therefore, the characteristic angular scale associated with this "horizon size" at the last scattering surface should be smaller than it would appear if the universe were flat, and in a closed universe it should be larger. Since late 1998 the BOOMERanG (Balloon Observations of Millimetric Extragalactic Radiation and Geophysics) experiment in Antarctica, as well as other balloon experiments in Canada and the U.S., has found definitive evidence of this angular signature's existence. Moreover, the fact that it corresponds to an angular scale of about one degree provides, for the very first time, a direct measurement of the geometry of the universe. And the universe does appear to be precisely flat.

Meanwhile researchers studying distant supernovae have provided the first direct, if tentative, evidence that the expansion of the universe is accelerating, a telltale sign of a cosmological constant with the same value implied by the other data [see "Surveying Spacetime with Supernovae," on page 22]. Observations of the microwave background and of supernovae illuminate two different aspects of cosmology. The microwave background reveals the geometry of the universe, which is sensitive to the total density of energy, in whatever form, where-

THE GEOMETRY OF THE UNIVERSE

By Martin A. Bucher and David N. Spergel

IF THE UNIVERSE had an "outside" and people could view it from that perspective, cosmology would be much easier. Lacking these gifts, astronomers must infer the basic shape of the universe from its geometric properties. Everyday experience indicates that space is Euclidean, or "flat," on small scales. Parallel lines never meet, triangles span 180 degrees, the circumference of a circle is $2\pi r$, and so on. But it would be wrong to assume that the universe is Euclidean on large scales, just as it would be wrong to conclude that the earth is flat just because a small patch of it looks flat.

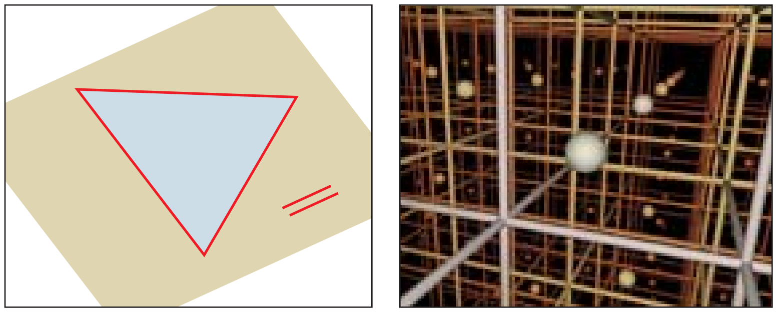
There are two other possible three-dimensional geometries consistent with the observations of cosmic homogeneity (the equivalence of all points in space) and isotropy (the equivalence of all directions). They are the spherical, or "closed," geometry and the hyperbolic, or "open," geometry. Both are characterized by a curvature length analogous to the earth's radius. If the curvature is positive, the geometry is spherical; if negative, hyperbolic. For distances much smaller than this length, all geometries look Euclidean.

In a spherical universe, as on the earth's surface, parallel lines eventually meet, triangles can span up to 540 degrees, and the circumference of a circle is smaller than $2\pi r$. Because the space curves back on itself, the spherical universe is finite. In a hyperbolic universe, parallel lines diverge, triangles have less than 180 degrees, and the circumference of a circle is larger than $2\pi r$. Such a universe, like Euclidean space, is infinite in size.

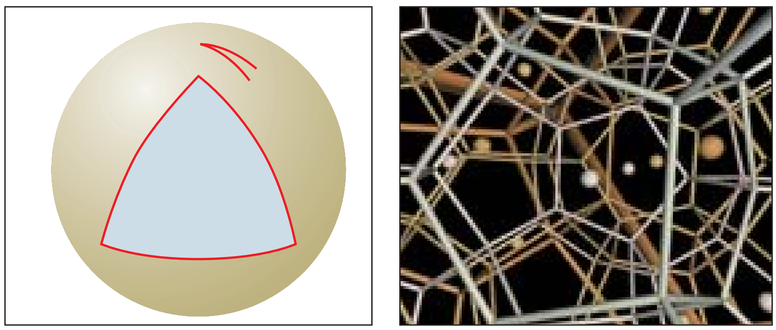
These three geometries have quite different effects on perspective (*see illustration at right*), which distort the appearance of features in the cosmic microwave background radiation. The largest ripples in the background have the same absolute size regardless of the process of inflation. If the universe is flat, the largest undulations would appear to be about one degree across. But if the universe is hyperbolic, the same features should appear to be only half that size, simply because of the geometric distortion of light rays.

Ground and balloon-borne observations suggest that the ripples are one degree across, which implies that the universe is nearly flat. The Microwave Anisotropy Probe, which is expected to return data soon, will make definitive measurements of these fluctations.

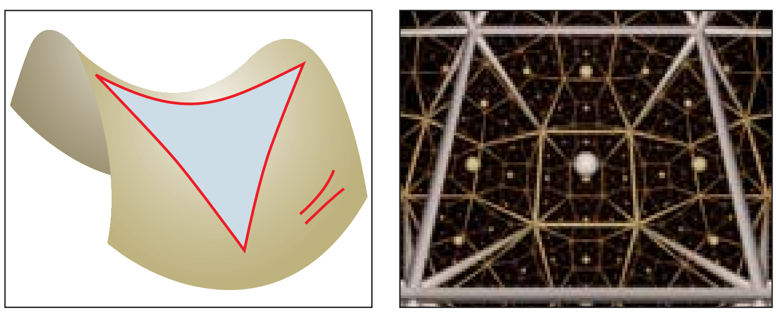
THREE GEOMETRIES are shown here from two different perspectives: a hypothetical outside view that ignores, for the sake of illustration, one of the spatial dimensions (*left column*) and an inside view that shows all three dimensions as well as a reference framework (*right column*). The outside view is useful for seeing the basic geometric rules. The inside view reveals the apparent sizes of objects (which, in these diagrams, are the same actual size) at different distances. Here objects and framework redden with distance.



Flat space obeys the familiar rules of Euclidean geometry. The angular size of identical spheres is inversely proportional to distance—the usual vanishing-point perspective taught in art class.



Spherical space has the geometric properties of a globe. With increasing distance, the spheres at first seem smaller. They reach a minimum apparent size and subsequently look larger. (Similarly, lines of longitude emanating from a pole separate, reach a maximum separation at the equator and then refocus onto the opposite pole.) This framework consists of dodecahedra.



Hyperbolic space has the geometry of a saddle. Angular size shrinks much more rapidly with distance than in Euclidean space. Because angles are more acute, five cubelike objects fit around each edge, rather than only four.

DON DIXON (*left column*); STUART LEVY *National Center for Supercomputing Applications* AND TAMARA MUNZNER *Stanford University* (*right column*)

Martin A. Bucher and David N. Spergel study the physics of the very early universe. Bucher is professor of applied mathematics and theoretical physics at the University of Cambridge. Spergel is professor of astrophysical sciences at Princeton University.

as the supernovae directly probe the expansion rate of the universe, which depends on the difference between the density of matter (which slows the expansion) and the cosmological constant (which can speed it up).

Together all these results suggest that the constant contributes between 50 and 75 percent of the energy needed to make the universe flat [*see illustration on page 35*]. Despite the preponderance of evidence, it is worth remembering the old saw that an astronomical theory whose predictions agree with all observations is probably wrong, if only because some of the measurements or some of the predictions are likely to be erroneous. Nevertheless, theorists are already scrambling to understand what 20 years ago would have been unthinkable: a cosmological constant greater than zero yet much smaller than current quantum theories predict. Some feat of fine-tuning must subtract virtual-particle energies to 123 decimal places but leave the 124th untouched—a precision seen nowhere else in nature.

One direction, explored by Steven Weinberg of the University of Texas at Austin and his colleagues, invokes the last resort of cosmologists, the anthropic principle. If the observed universe is merely one of an infinity of disconnected universes—each of which might have slightly different constants of nature, as suggested by some incarnations of inflationary theory combined with emerging ideas of quantum gravity—then physicists can hope to estimate the magnitude of the cosmological constant by asking in which universes intelligent life is likely to evolve. Weinberg and others have arrived at a result that is compatible with the apparent magnitude of the cosmological constant.

Most theorists, however, do not find these notions convincing, as they imply that there is no reason for the constant to take on a particular value; it just does. Although that argument may be true, physicists have not yet exhausted the other possibilities, which might allow the constant to be constrained by fundamental theory rather than by accidents of history.

Another direction of research follows in a tradition established by Dirac. He argued that there is one measured large number in the universe—its age (or, equivalently, its size). If certain physical quantities were changing over time, they might naturally be either very large or very small today [see "P.A.M. Dirac and the Beauty of Physics," by R. Corby Hovis and Helge Kragh; SCIENTIFIC AMERICAN, May 1993]. The cosmological constant could be one example. It might not, in fact, be constant. After all, if the cosmological constant is fixed and nonzero, we are living at the first and only time in the cosmic history when the density of matter, which decreases as the universe expands, is comparable to the energy stored in empty space. Why the coincidence? Several groups have instead imagined that some form of cosmic energy mimics a cosmological constant but varies with time.

This concept was explored by P. James E. Peebles and Bharat V. Ratra of Princeton University more than a decade ago. Motivated by the new supernova findings, other groups have resurrected the idea. Some have drawn on emerging concepts from string theory. Robert R. Caldwell of Dartmouth College and Paul J. Steinhardt of Princeton have reproposed the term "quintessence" to describe this variable energy [see "The Quintessential Universe," on page 40]. It is one measure of the theoretical conundrum that the dark matter that originally deserved this term now seems almost mundane by comparison. As much as I like the word, none of the theoretical ideas for quintessence seems compelling. Each is ad hoc. The enormity of the cosmological-constant problem remains.

How will cosmologists know for certain whether they have to reconcile themselves to this theoretically perplexing universe? New measurements of the microwave background and of galaxy evolution, the continued analysis of distant supernovae and measurements of gravitational lensing of distant quasars should be able to pin down the cosmological constant over the next few years. One thing is already certain. The standard cosmology of the 1980s, postulating a flat universe dominated by matter, is dead. The universe appears to be filled with an energy of unknown origin. This will require a dramatic new understanding of physics. Put another way, "nothing" could not possibly be more interesting. SA

Summary of Inferred Values of Cosmic Matter Density

MEASUREMENTS of the contribution to Ω from matter are in rough concordance. Most astronomers now accept that matter alone cannot make Ω equal to 1. But other forms of energy, such as the cosmological constant, may also pitch in. —*L.M.K.*

OBSERVATION	Ω_{MATTER}
Age of universe	<1
Density of protons and neutrons	0.3–0.6
Galaxy clustering	0.3–0.4
Galaxy evolution	0.3–0.5
Cosmic microwave background radiation	≲1
Supernovae type Ia	0.2–0.5

MORE TO EXPLORE

Dreams of a Final Theory. Steven Weinberg. Pantheon Books, 1992.

Principles of Physical Cosmology. P. James E. Peebles. Princeton University Press, 1993.

Before the Beginning: Our Universe and Others. Martin Rees. Addison-Wesley, 1997.

The Age of Globular Clusters in Light of Hipparcos: Resolving the Age Problem? Brian Chaboyer, Pierre Demarque, Peter J. Kernan and Lawrence M. Krauss in *Astrophysical Journal,* Vol. 494, No. 1, pages 96–110; February 10, 1998. Preprint available at xxx.lanl.gov/abs/astro-ph/9706128

The End of the Age Problem, and the Case for a Cosmological Constant Revisited. Lawrence M. Krauss in *Astrophysical Journal,* Vol. 501, No. 2, pages 461–466; July 10, 1998. Preprint available at xxx.lanl.gov/abs/astro-ph/9706227

Living with Lambda. J. D. Cohn. Preprint available at xxx.lanl.gov/abs/astro-ph/9807128

Quintessence. Lawrence M. Krauss. Basic Books, 2001.

the quintessential UNIVERSE

The universe has recently been commandeered by an invisible energy field, which is causing its expansion to accelerate outward

BY JEREMIAH P. OSTRIKER AND PAUL J. STEINHARDT

DON DIXON

Updated from the January 2001 issue

Is it all over but the shouting?

Is the cosmos understood aside from minor details? A few years ago it certainly seemed that way. After a century of vigorous debate, scientists had reached a broad consensus about the basic history of the universe. It all began with gas and radiation of unimaginably high temperature and density. For 15 billion years, it has been expanding and cooling. Galaxies and other complex structures have grown from microscopic seeds—quantum fluctuations—that were stretched to cosmic size by a brief period of "inflation." We had also learned that only a small fraction of matter is composed of the normal chemical elements of our everyday experience. The majority consists of so-called dark matter, primarily exotic elementary particles that do not interact with light. Plenty of mysteries remained, but at least we had sorted out the big picture.

Or so we thought. It turns out that we have been missing most of the story. Over the past five years or so, observations have convinced cosmologists that the chemical elements and the dark matter, combined, amount to less than half the content of the universe. The bulk is a ubiquitous dark energy with a strange and remarkable feature: its gravity does not attract. It repels. Whereas gravity pulls the chemical elements and dark matter into stars and galaxies, it pushes the dark energy into a nearly uniform haze that permeates space. The universe is a battleground between the two tendencies, and repulsive gravity is winning. It is gradually overwhelming the attractive force of ordinary matter—causing the universe to accelerate to ever larger rates of expansion, perhaps leading to a new runaway inflationary phase and a totally different future for the universe than most cosmologists envisioned a decade ago.

Until recently, cosmologists have focused simply on proving the existence of dark energy. Having made a convincing case, they are now turning their attention to a deeper problem: Where does the energy come from? The best-known possibility is that the energy is inherent in the fabric of space. Even if a volume of space were utterly empty—without a bit of matter and radiation—it would still contain this energy. Such energy is a venerable notion that dates back to Albert Einstein and his attempt in 1917 to construct a static model of the universe. Like many leading scientists over the centuries, including Isaac Newton, Einstein believed that the universe is unchanging, neither contracting nor expanding. To coax stagnation from his general theory of relativity, he had to introduce vacuum energy, or, in his terminology, a cosmological constant. He adjusted the value of the constant so that its gravitational repulsion would exactly counterbalance the gravitational attraction of matter.

Later, when astronomers established that the universe is expanding, Einstein regretted his delicately tuned artifice, calling it his greatest blunder. But perhaps his judgment was too hasty. If the cosmological constant had a slightly larger value than Einstein proposed, its repulsion would exceed the attraction of matter, and cosmic expansion would accelerate.

Many cosmologists, though, are now leaning toward a different idea, known as quintessence. The translation is "fifth element," an allusion to ancient Greek philosophy, which suggested that the universe is composed of earth, air, fire and water, plus an ephemeral substance that prevents the moon and planets from falling to the center of the celestial sphere. Four years ago Robert R. Caldwell, Rahul Dave and one of us (Steinhardt), all then at the University of Pennsylvania, reintroduced the term to refer to a dynamical quantum field, not unlike an electrical or magnetic field, that gravitationally repels.

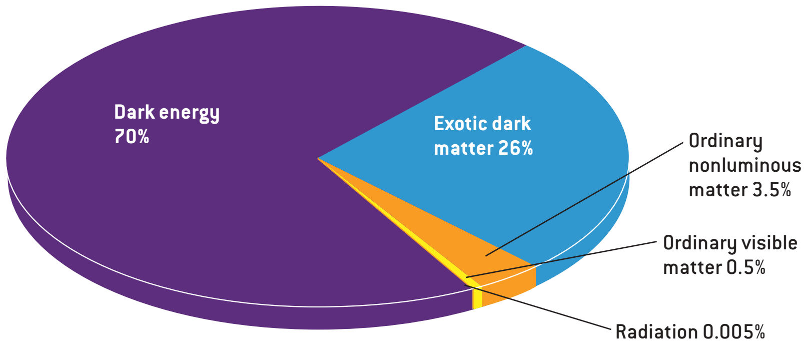
Dark energy differs from dark matter in one major respect: it must be gravitationally repulsive.

The dynamism is what cosmologists find so appealing about quintessence. The biggest challenge for any theory of dark energy is to explain the inferred amount of the stuff—not so much that it would have interfered with the formation of stars and galaxies in the early universe but just enough that its effect can now be felt. Vacuum energy is completely inert, maintaining the same density for all time. Consequently, to explain the amount of dark energy today, the value of the cosmological constant would have to be fine-tuned at the creation of the universe to have the proper value—which makes it sound rather like a fudge factor. In contrast, quintessence interacts with matter and evolves with time, so it might naturally adjust itself to reach the observed value today.

JANA BRENNING; SOURCE: PAUL J. STEINHARDT

RECIPE FOR THE UNIVERSE

The main ingredient of the universe is dark energy, which consists of either the cosmological constant or the quantum field known as quintessence. The other ingredients are dark matter (composed of exotic elementary particles), ordinary matter (both nonluminous and visible), and a trace amount of radiation.



Percentages do not add up to 100 because of rounding.

JANA BRENNING

Two Thirds of Reality

DISTINGUISHING between these two options is critically important for physics. Particle physicists have depended on high-energy accelerators to discover new forms of energy and matter. Now the cosmos has revealed an unanticipated type of energy, too thinly spread and too weakly interacting for accelerators to probe. Determining whether the energy is inert or dynamical may be crucial to formulating a fundamental theory of nature. Particle physicists are discovering that they must keep a close eye on developments in the heavens as well as those in the accelerator laboratory.

The case for dark energy has been building brick by brick for nearly a decade. The first brick was a thorough census of all matter in galaxies and galaxy clusters using a variety of optical, x-ray and radio techniques. The unequivocal conclusion was that the total mass in chemical elements and dark matter accounts for only about one third of the quantity that most theorists expected—the so-called critical density.

Many cosmologists took this as a sign that the theorists were wrong. In that case, we would be living in an ever expanding universe where space is curved hyperbolically, like the horn on a trumpet. But this interpretation has been put to rest by measurements of hot and cold spots in the cosmic microwave background radiation, whose distribution has shown that space is flat and that the total energy density equals the critical density. Putting the two observations together, simple arithmetic dictates the necessity for an additional energy component to make up the missing two thirds of the energy density.

Whatever it is, the new component must be dark, neither absorbing nor emitting light, or else it would have been noticed long ago. In that way, it resembles dark matter. But the new component—called dark energy—differs from dark matter in one major respect: it must be gravitationally repulsive. Otherwise it would be pulled into galaxies and clusters, where it would affect the motion of visible matter. No such influence is seen. Moreover, gravitational repulsion resolves the "age crisis" that plagued cosmology in the 1990s. If one takes the current measurements of the expansion rate and assumes that the expansion has been decelerating, the age of the universe is less than 12 billion years.

Yet evidence suggests that some stars in our galaxy are 15 billion years old. By causing the expansion rate of the universe to accelerate, repulsion brings the inferred age of the cosmos into agreement with the observed age of celestial bodies [see "Cosmological Antigravity," on page 30].

The potential flaw in the argument used to be that gravitational repulsion should cause the expansion to accelerate, which had not been observed. Then, in 1998, the last brick fell into place. Two independent groups used measurements of distant supernovae to detect a change in the expansion rate. Both groups concluded that the universe is accelerating and at just the pace predicted [see "Surveying Spacetime with Supernovae," on page 22].

All these observations boil down to three numbers: the average density of matter (both ordinary and dark), the average density of dark energy, and the curvature of space. Einstein's equations dictate that the three numbers add up to the critical density. The possible combinations of the numbers can be succinctly represented on a triangular plot [*see illustration on page 49*]. The three distinct sets of observations—matter census, cosmic microwave background, and supernovae—correspond to strips inside the triangle. Remarkably, the three strips overlap at the same position, which makes a compelling case for dark energy.

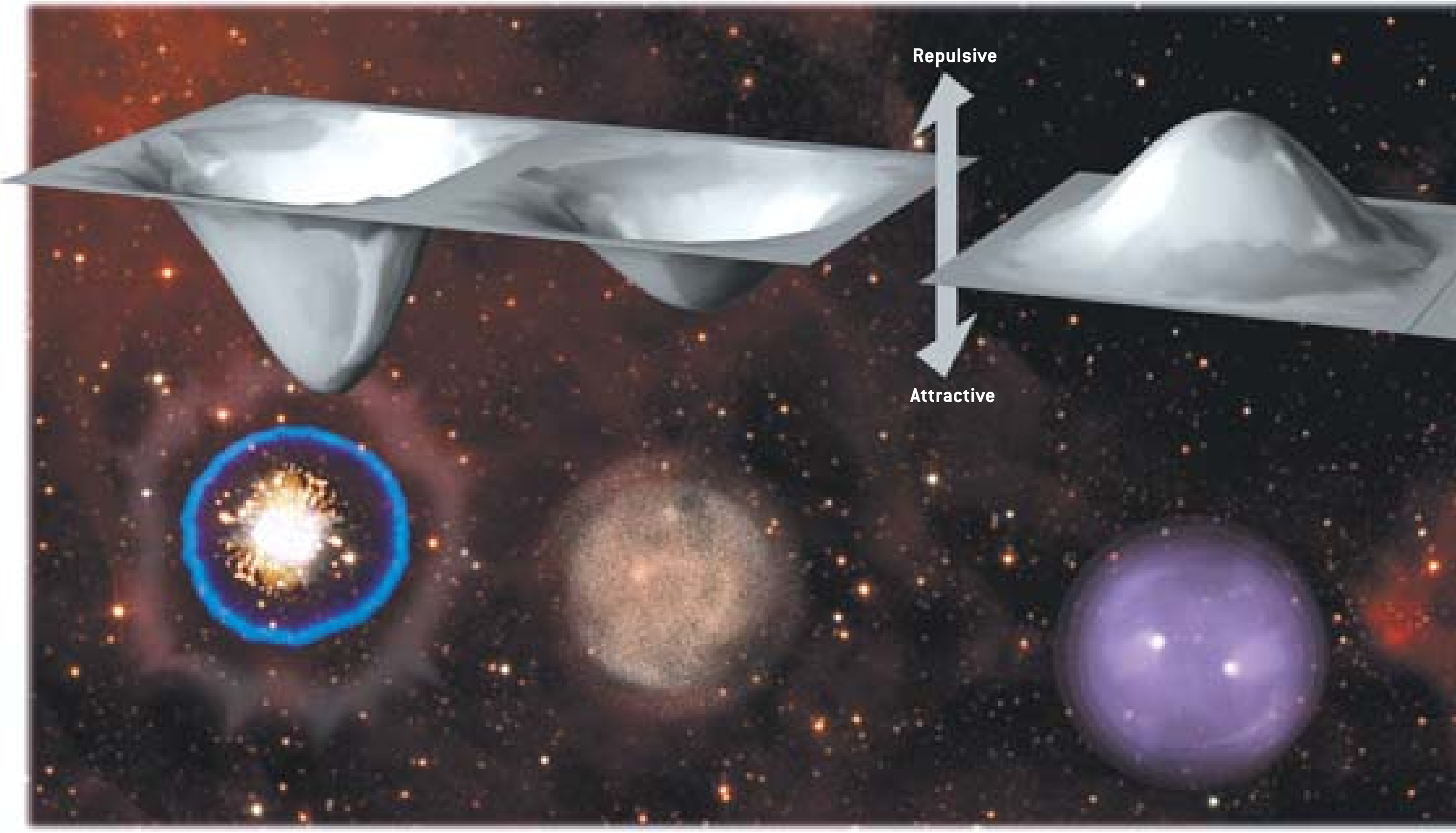
From Implosion to Explosion

OUR EVERYDAY EXPERIENCE is with ordinary matter, which is gravitationally attractive, so it is difficult to envisage how dark energy could gravitationally repel. The key feature is that its pressure is negative. In Newton's law of gravity, pressure plays no role; the strength of gravity depends only on mass. In Einstein's law of gravity, however, the strength of gravity depends not just on mass but also on other forms of energy and on pressure. In this way, pressure has two effects: direct (caused by the action of the pressure on surrounding material) and indirect (caused by the gravitation that the pressure creates).

The sign of the gravitational force is

THE AUTHORS

JEREMIAH P. OSTRIKER and *PAUL J. STEINHARDT*, both professors at Princeton University, have been collaborating for the past seven years. Their prediction of accelerating expansion in 1995 anticipated groundbreaking supernova results by several years. Ostriker was one of the first scientists to appreciate the prevalence of dark matter and the importance of hot intergalactic gas. In 2000 he won the U.S. National Medal of Science. Steinhardt was one of the originators of the theory of inflation and the concept of quasicrystals. He reintroduced the term "quintessence" after his youngest son, Will, and daughter Cindy picked it over several alternatives.



Radiation

Ordinary matter

Quintessence (moderately negative pressure)

THE POWER OF POSITIVE (AND NEGATIVE) THINKING

Whether a lump of energy exerts a gravitationally attractive or repulsive force depends on its pressure. If the pressure is zero or positive, as it is for radiation or ordinary matter, gravity is attractive. (The downward dimples represent the potential energy wells.) Radiation has greater pressure, so its gravity is more attractive. For quintessence, the pressure is negative and gravity is repulsive (the dimples become hills).

DON FOLEY; SOURCE: ROBERT R. CALDWELL *Dartmouth College* AND PAUL J. STEINHARDT

determined by the algebraic combination of the total energy density plus three times the pressure. If the pressure is positive, as it is for radiation, ordinary matter and dark matter, then the combination is positive and gravitation is attractive. If the pressure is sufficiently negative, the combination is negative and gravitation is repulsive. To put it quantitatively, cosmologists consider the ratio of pressure to energy density, known as the equation of state, or w. For an ordinary gas, w is positive and proportional to the temperature. But for certain systems, w can be negative. If it drops below $-1/3$, gravity becomes repulsive.

Vacuum energy meets this condition (provided its density is positive). This is a consequence of the law of conservation of energy, according to which energy cannot be destroyed. Mathematically the law can be rephrased to state that the rate of change of energy density is proportional to $w + 1$. For vacuum energy—whose density, by definition, never changes—this sum must be zero. In other words, w must equal precisely -1. So the pressure must be negative.

What does it mean to have negative pressure? Most hot gases have positive pressure; the kinetic energy of the atoms and radiation pushes outward on the container. Note that the direct effect of positive pressure—to push—is the opposite of its gravitational effect—to pull. But one can imagine an interaction among atoms that overcomes the kinetic energy and causes the gas to implode. The implosive gas has negative pressure. A balloon of this gas would collapse inward because the outside pressure (zero or positive) would exceed the inside pressure (negative). Curiously, the direct effect of negative pressure—implosion—can be the opposite of its gravitational effect—repulsion.

The gravitational effect is tiny for a balloon. But now imagine filling all of space with the implosive gas. Then there is no bounding surface and no external pressure. The gas still has negative pressure, but it has nothing to push against, so it exerts no direct effect. It has only the gravitational effect—namely, repulsion. The repulsion stretches space, increasing its volume and, in turn, the amount of vacuum energy. The tendency to stretch is therefore self-reinforcing. The universe expands at an accelerating pace. The growing vacuum energy comes at the expense of the gravitational field.

These concepts may sound strange, and even Einstein found them hard to swallow. He viewed the static universe, the original motivation for vacuum ener-

Quintessence
(highly negative pressure)

gy, as an unfortunate error that ought to be dismissed. But the cosmological constant, once introduced, would not fade away. Theorists soon realized that quantum fields possess a finite amount of vacuum energy, a manifestation of quantum fluctuations that conjure up pairs of virtual particles from scratch. An estimate of the total vacuum energy produced by all known fields predicts a huge amount—120 orders of magnitude more than the energy density in all other matter. That is, though it is hard to picture, the evanescent virtual particles should contribute a positive, constant energy density, which would imply negative pressure. But if this estimate were true, an acceleration of epic proportions would rip apart atoms, stars and galaxies. Clearly, the estimate is wrong. One of the major goals of unified theories of gravity has been to learn why.

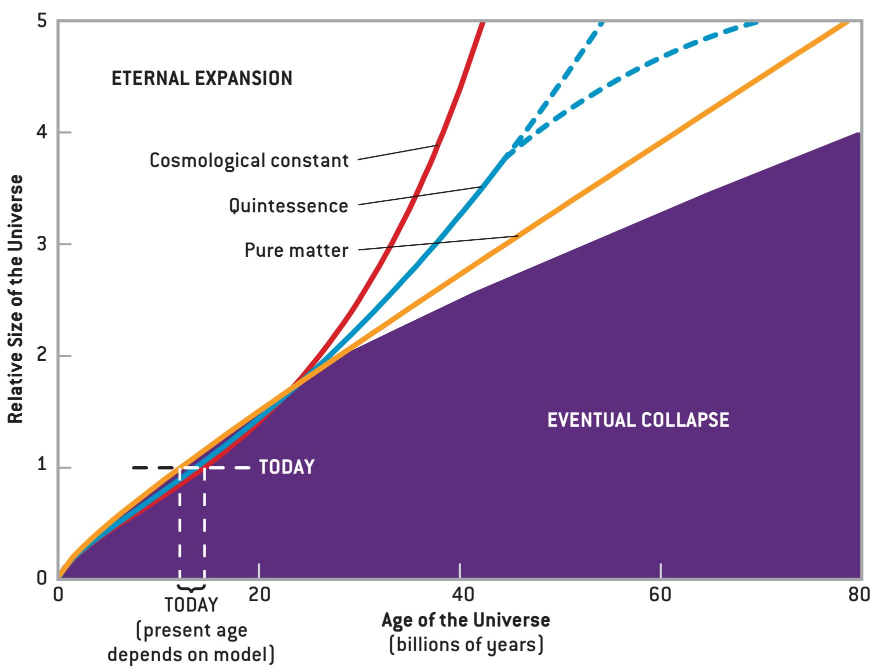
One proposal is that some heretofore undiscovered symmetry in fundamental physics results in a cancellation of large effects, zeroing out the vacuum energy. For example, quantum fluctuations of virtual pairs of particles contribute positive energy for particles with half-integer spin (like quarks and electrons) but negative energy for particles with integer spin (like photons). In standard theories, the cancellation is inexact, leaving behind an unacceptably large energy density. But physicists have been exploring models with so-called supersymmetry, a relation between the two particle types that can lead to a precise cancellation. A serious flaw, though, is that supersymmetry would be valid only at very high energies. Theorists are working on a way of preserving the perfect cancellation even at lower energies.

Another thought is that the vacuum energy is not exactly nullified after all. Perhaps there is a cancellation mechanism that is slightly imperfect. Instead of making the cosmological constant exactly zero, the mechanism cancels only to 120 decimal places. Then the vacuum energy could constitute the missing two thirds of the universe. That seems bizarre, though. What mechanism could possibly work with such precision? Although the dark energy represents a huge amount of mass, it is spread so thinly that its energy is less than four electron volts per cubic millimeter—which, to a particle physicist, is unimaginably low. The weakest known force in nature involves an energy density 1,050 times greater.

Extrapolating back in time, vacuum energy gets even more paradoxical. Today matter and dark energy have comparable average densities. But billions of years ago, when they came into being, our universe was the size of a grapefruit, so matter was 100 orders of magnitude denser. The cosmological constant, however, would have had the same value as it does now. In other words, for every 10,100 parts matter, physical processes would have created one part vacuum energy—a degree of exactitude that may be reasonable in a mathematical idealization but that seems ludicrous to expect from the real world. This need for almost supernatural fine-tuning is the principal motivation for considering alternatives to the cosmological constant.

Fieldwork

FORTUNATELY, vacuum energy is not the only way to generate negative pres-



GROWING PAINS

The universe expands at different rates depending on which form of energy predominates. Matter causes the growth to decelerate, whereas the cosmological constant causes it to accelerate. Quintessence is in the middle: it forces the expansion to accelerate, but less rapidly. Eventually the acceleration may or may not switch off (*dashed lines*).

JANA BRENNING; SOURCE: ROBERT R. CALDWELL *Dartmouth College* AND PAUL J. STEINHARDT sure. Another means is an energy source that, unlike vacuum energy, varies in space and time—a realm of possibilities that goes under the rubric of quintessence. For quintessence, w has no fixed value, but it must be less than $-1/3$ for gravity to be repulsive.

Quintessence may take many forms. The simplest models propose a quantum field whose energy is varying so slowly that it looks, at first glance, like a constant vacuum energy. The idea is borrowed from inflationary cosmology, in which a cosmic field known as the inflaton drives expansion in the very early universe using the same mechanism. The key difference is that quintessence is much weaker than the inflaton. This hypothesis was first explored a decade ago by Christof Wetterich of the University of Heidelberg and by Bharat Ratra, now at Kansas State University, and P. James E. Peebles of Princeton University.

In quantum theory, physical processes can be described in terms either of fields or of particles. But because quintessence has such a low energy density and varies so gradually, a particle of quintessence would be inconceivably lightweight and large—the size of a supercluster of galaxies. So the field description is rather more useful. Conceptually, a field is a continuous distribution of energy that assigns to each point in space a numerical value known as the field strength. The energy embodied by the field has a kinetic component, which depends on the time variation of the field strength, and a potential component, which depends only on the value of the field strength. As the field changes, the balance of kinetic and potential energy shifts.

For vacuum energy, recall that the negative pressure was the direct result of the conservation of energy, which dictates that any variation in energy density is proportional to the sum of the energy density (a positive number) and the pressure. For vacuum energy, the change is zero, so the pressure must be negative. For quintessence, the change is gradual enough that the pressure must still be negative, though somewhat less so. This condition corresponds to having more potential energy than kinetic energy.

Because its pressure is less negative, quintessence does not accelerate the universe as strongly as vacuum energy does. Ultimately this will be how observers decide between the two. If anything, quintessence is more consistent with the available data, but for now the distinction is not statistically significant. Another difference is that, unlike vacuum energy, the quintessence field may undergo all kinds of complex evolution. The value of w may be positive, then negative, then positive again. It may have different values in different places. Although the nonuniformity is thought to be small, it may be detectable by studying the cosmic microwave background radiation.

A further difference is that quintessence can be perturbed. Waves will propagate through it just as sound waves can pass through the air. In the jargon, quintessence is "soft." Einstein's cosmological constant is, in contrast, stiff—it cannot be pushed around. This raises an interesting issue. Every known form of energy is soft to some degree. Perhaps stiffness is an idealization that cannot exist in reality, in which case the cosmological constant is an impossibility. Quintessence with w near -1 may be the closest reasonable approximation.

Quintessence might fuel a cyclic model in which the hot, homogeneous universe is made and remade eternally.

Quintessence on the Brane

SAYING THAT quintessence is a field is just the first step in explaining it. Where would such a strange field come from? Particle physicists have explanations for phenomena from the structure of atoms to the origin of mass, but quintessence is something of an orphan. Modern theories of elementary particles include many kinds of fields that might have the requisite behavior, but not enough is known about their kinetic and potential energy to say which, if any, could produce negative pressure today.

An exotic possibility is that quintessence springs from the physics of extra dimensions. Over the past few decades, theorists have been exploring string theory, which may combine general relativity and quantum mechanics in a unified theory of fundamental forces. An important feature of string models is that they predict 10 dimensions. Four of these are our familiar three spatial dimensions, plus time. The remaining six must be hidden. In some formulations, they are curled up like a ball whose radius is too small to be detected (at least with present instruments). An alternative idea is found in a recent extension of string theory, known as M-theory, which adds an 11th dimension: ordinary matter is confined to two three-dimensional surfaces, known as branes (short for membranes), separated by a microscopic gap along the 11th dimension [see "The Universe's Unseen Dimensions," on page 66].

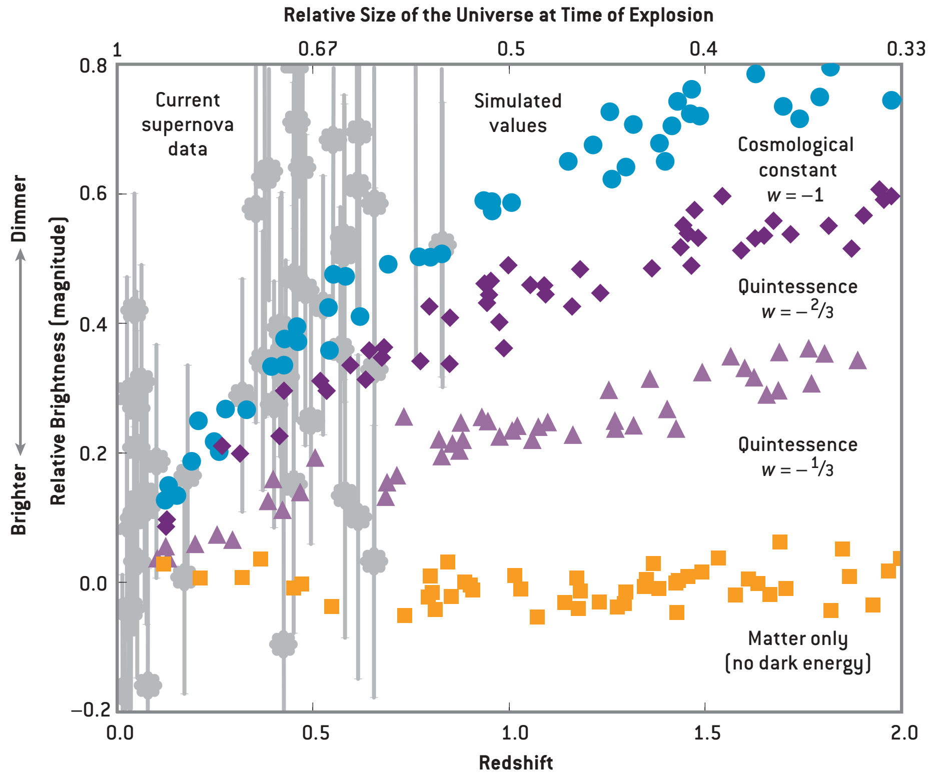
We are unable to see the extra dimensions, but if they exist, we should be able to perceive them indirectly. In fact, the presence of curled-up dimensions or nearby branes would act just like a field. The numerical value that the field assigns to each point in space could correspond to the radius or gap distance. If the radius or gap changes slowly as the universe expands, it could exactly mimic the hypothetical quintessence field.

Whatever the origin of quintessence, its dynamism could solve the thorny problem of fine-tuning. One way to look at this issue is to ask why cosmic acceleration has begun at this particular moment in cosmic history. Created when the universe was 10^{-35} second old, dark energy must have remained in the shadows for nearly 10 billion years—a factor of more than 10^{50} in age. Only then, the data suggest, did it overtake matter and cause the universe to begin accelerating. Is it not a coincidence that, just when thinking beings evolved, the universe suddenly shifted into overdrive? Somehow the fates of matter and dark energy seem to be intertwined. But how?

If the dark energy is vacuum energy, the coincidence is almost impossible to account for. Some researchers, including Martin Rees of the University of Cambridge and Steven Weinberg of the University of Texas at Austin, have pursued an anthropic explanation. Perhaps our universe is just one among a multitude of universes, in each of which the vacuum energies takes on a different value. Universes with vacuum energies much greater than four electron volts per cubic millimeter might be more common, but they expand too rapidly to form stars, planets or life. Universes with much smaller values might be very rare. Our universe would have the optimal value. Only in this "best of all worlds" could there exist intelligent beings capable of contemplating the nature of the universe. But physicists disagree whether the anthropic argument constitutes an acceptable explanation [see "Exploring Our Universe and Others," on page 82].

A more satisfying answer, which could involve a form of quintessence known as a tracker field, was studied by Ratra and Peebles and by Steinhardt and Ivaylo Zlatev and Limin Wang, then at the University of Pennsylvania. The equations that describe tracker fields have classical attractor behavior like that found in some chaotic systems. In such systems, motion converges to the same result for a wide range of initial conditions. A marble put into an empty bathtub, for example, ultimately falls into the drain whatever its starting place.

Similarly, the initial energy density of the tracker field does not have to be tuned to a certain value, because the field rapidly adjusts itself to that value. It locks into a track on which its energy density remains a nearly constant fraction of the



SEEING WILL BE BELIEVING

Supernova data may be one way to decide between quintessence and the cosmological constant. The latter makes the universe speed up faster, so supernovae at a given redshift would be farther away and hence dimmer. Existing telescopes (*data shown in gray*) cannot tell the two cases apart, but the proposed Supernova Acceleration Probe should be able to. The supernova magnitudes predicted by four models are shown in different colors.

JANA BRENNING; SOURCE: IRIT MAOR AND RAM BRUSTEIN *Ben-Gurion University*

density of radiation and matter. In this sense, quintessence imitates matter and radiation, even though its composition is wholly different. The mimicking occurs because the radiation and matter density determine the cosmic expansion rate, which, in turn, controls the rate at which the quintessence density changes. On closer inspection, one finds that the fraction is slowly growing. Only after many millions or billions of years does quintessence catch up.

So why did quintessence catch up when it did? Cosmic acceleration could just as easily have begun in the distant past or in the far future, depending on the choices of constants in the tracker-field theory. This brings us back to the coincidence. But perhaps some event in the relatively recent past unleashed the acceleration. Steinhardt, along with Christian Armendáriz Picon, now at the University of Chicago, and Viatcheslav Mukhanov of Ludwig Maximilians University in Munich, has proposed one such recent event: the transition from radiation domination to matter domination.

According to the big bang theory, the energy of the universe used to reside mainly in radiation. As the universe cooled, however, the radiation lost energy faster than ordinary matter did. By the time the universe was a few tens of thousands of years old—a relatively short time ago in logarithmic terms—the energy balance had shifted in favor of matter. This change marked the beginning of the matter-dominated epoch of which we are the beneficiaries. Only then could gravity begin to pull matter together to form galaxies and larger-scale structures. At the same time, the expansion rate of the universe underwent a change.

In a variation on the tracker models, this transformation triggered a series of events that led to cosmic acceleration today. Throughout most of the history of the universe, quintessence tracked the radiation energy, remaining an insignificant component of the cosmos. But when the universe became matter-dominated, the change in the expansion rate jolted quintessence out of its copycat behavior. Instead of tracking radiation or even matter, the pressure of quintessence switched to a negative value. Its density stayed nearly fixed and ultimately overtook the decreasing matter density. In this picture, the fact that thinking beings and cosmic acceleration came into existence at nearly the same time is not a coincidence. Both the formation of stars and planets necessary to support life and the transformation of quintessence into a negative-pressure component were triggered by the onset of matter domination.

Looking to the Future

IN THE SHORT TERM, the focus of cosmologists will be to detect the existence of quintessence. It has observable consequences. Because its value of *w* differs from that of vacuum energy, it produces a different rate of cosmic acceleration. More precise measurements of supernovae over a longer span of distances may separate the two cases. Astronomers have proposed two new observatories—the orbiting Supernova Acceleration Probe and the Earth-based Large-Aperture Synoptic Survey Telescope—to resolve the issue. Differences in acceleration rate also produce small differences in the angular size of hot and cold spots in the cosmic microwave background radiation, as the Microwave Anisotropy Probe (MAP) and Planck spacecraft should be able to detect.

Other tests measure how the number of galaxies varies with increasing redshift to infer how the expansion rate of the universe has changed with time. A ground-based project known as the Deep Extragalactic Evolutionary Probe will look for this effect.

Over the longer term, all of us will be left to ponder the profound implications of these revolutionary discoveries. They lead to a sobering new interpretation of our place in cosmic history. In the beginning (or at least at the earliest time for which we have any clue), there was inflation, an extended period of accelerated expansion during the first instants after the big bang. Space back then was nearly devoid of matter, and a quintessence-like quantum field with negative pressure held sway. During that period, the universe expanded by a greater factor than it has during the 15 billion years since inflation ended. At the end of inflation, the field decayed to a hot gas of quarks, gluons, electrons, light, and dark energy.

For thousands of years, space was so thick with radiation that atoms, let alone larger structures, could not form. Then matter took control. The next stage—our epoch—has been one of steady cooling, condensation and the evolution of intricate structure of ever increasing size. But this period is coming to an end. Cosmic acceleration is back. The universe as we know it, with shining stars, galaxies and clusters, appears to have been a brief interlude. As acceleration takes hold over the next tens of billions of years, the matter and energy in the universe will become more and more diluted and space will stretch too rapidly to enable new structures to form. Living things will find the cosmos increasingly hostile [see "The Fate of Life in the Universe," on page 50]. If the acceleration is caused by vacuum energy, then the cosmic story is complete: the planets, stars and galaxies we see today are the pinnacle of cosmic evolution.

But if the acceleration is caused by quintessence, the ending has yet to be written. The universe might accelerate forever, or the quintessence could decay into new forms of matter and radiation, repopulating the universe. Because the dark energy density is so small, one might suppose that the material derived from its decay would have too little energy to do anything of interest. Under some circumstances, however, quintessence could decay through the nucleation of bubbles. The bubble interior would be a void, but the bubble wall would be the site of vigorous activity. As the wall moved outward, it would sweep up all the energy derived from the decay of quintessence. Occasionally, two bubbles would collide in a fantastic fireworks display. In the process, massive particles such as protons and neutrons might arise—perhaps stars and planets.

To future inhabitants, the universe would look highly inhomogeneous, with life confined to distant islands surrounded by vast voids. Would they ever figure out that their origin was the homogeneous and isotropic universe we see about us today? Would they ever know that the universe had once been alive and then died, only to be given a second chance?

Special survey instruments, plus new tests, will tell us which future is ours.

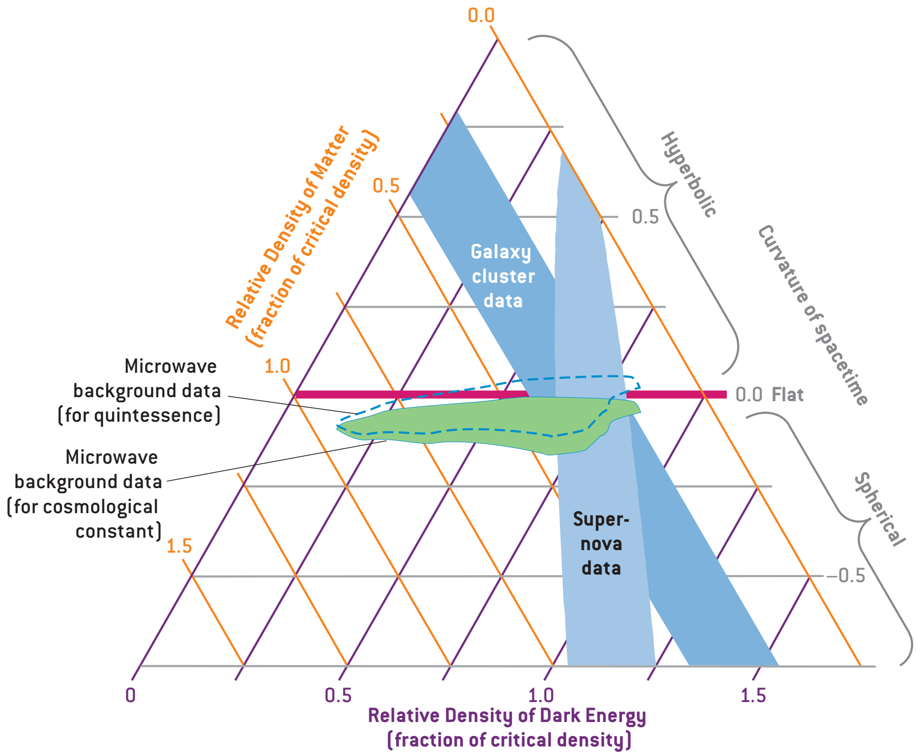
Or perhaps a more radical revision of cosmic history is in store. Inspired by the recent observations of cosmic acceleration, Steinhardt and Neil Turok of the University of Cambridge have proposed a "cyclical universe" model in which quintessence is center stage and inflation is excised altogether. In this picture, space and time exist forever. The universe undergoes an endless sequence of cycles in which it contracts in a big crunch and reemerges in an expanding big bang, with trillions of years of evolution in between. During the first 15 billion years of each cycle, the universe is dominated by radiation and matter, and as the universe cools, galaxies and stars form. Then, just as we are seeing today, quintessence initiates an extended period of accelerated expansion that empties the universe of the matter and entropy created in the previous cycle. Quintessence plays the essential role of making the universe homogeneous and at the same time flat-tens the spatial geometry—two of the functions that are usually attributed to inflation.

In addition, fluctuations in the quintessence field eventually form the seeds for galaxy formation after the bang, the third function played by inflation. As the quintessence field evolves, its density and pressure change until the field ceases to cause acceleration and instead initiates a period of contraction. At the crunch, some of the energy of the quintessence field is converted into the matter and radiation that fuel the bang and a new period of expansion, cooling and structure formation. Notably, the temperature and density rise to a large but finite density. So the model also avoids the infinities of the conventional big bang view. The hot, homogeneous universe is made and remade eternally.

The cyclic scenario has a natural interpretation in terms of the superstring picture of branes and extra dimensions. The cycles can be described as an infinite, periodic sequence of collisions between branes. Each collision creates a bang in which new matter and radiation are created. The radiation and matter cause the branes to stretch—the usual period of cosmic expansion. Yet there is also a force between the branes that contributes a positive potential energy to the universe when the branes are far apart. In this scenario, quintessence is simply this potential energy. After 15 billion years of expansion, the interbrane potential energy dominates the universe and a period of cosmic acceleration begins. The branes stretch sufficiently to dilute the density of matter and radiation and flatten any curvature or wrinkles in the branes' surfaces.

The branes move together slowly, but as they approach, the potential energy eventually decreases from a positive to a negative value. The quintessence field now causes the stretching to stop and the branes to speed toward collision. The collision and bounce correspond to the reversal from contraction to expansion. Yet only the extra-dimensions component collapses and reappears. The usual three-dimensions component remains infinite. Hence, the density of matter on the branes remains small and dilute even at the crunch. When the two branes bounce apart, the potential energy is restored to its original value and quintessence is recreated in preparation for the next cycle.

Experiments may soon give us some idea which future is ours. We trust that improved accuracy of the classic cosmological tests, plus specially designed survey instruments and some new tests (possibly using gravitational lensing), will make this possible. Will it be the dead end of vacuum energy or the untapped potential of quintessence? Ultimately the answer depends on whether quintessence has a place in the basic workings of nature—the realm, perhaps, of string theory. Our place in cosmic history hinges on the interplay between the science of the very big and that of the very small. SA



COSMIC TRIANGLE

In this graph of cosmological observations, the axes represent possible values of three key characteristics of the universe. If the universe is flat, as inflationary theory suggests, the different types of observations (*colored areas*) and the zero-curvature line (*red line*) should overlap. At present, the microwave background data produce a slightly better overlap if dark energy consists of quintessence (*dashed outline*) rather than the cosmological constant (*green area*).

JANA BRENNING; SOURCE: PAUL J. STEINHARDT

MORE TO EXPLORE

Cosmological Imprint of an Energy Component with General Equation of State. Robert R. Caldwell, Rahul Dave and Paul J. Steinhardt in *Physical Review Letters,* Vol. 80, No. 8, pages 1582–1585; February 23, 1998. Available at **xxx.lanl.gov/abs/astro-ph/9708069**

Cosmic Concordance and Quintessence. Limin Wang, R. R. Caldwell, J. P. Ostriker and Paul J. Steinhardt in *Astrophysical Journal,* Vol. 530, No. 1, Part 1, pages 17–35; February 10, 2000. Available at **astro-ph/9901388**

Dynamical Solution to the Problem of a Small Cosmological Constant and Late-Time Cosmic Acceleration. C. Armendáriz Picon, V. Mukhanov and Paul J. Steinhardt in *Physical Review Letters,* Vol. 85, No. 21, pages 4438–4441; November 20, 2000. Available at **astro-ph/0004134**

Why Cosmologists Believe the Universe Is Accelerating. Michael S. Turner in *Type Ia Supernovae: Theory and Cosmology.* Edited by Jens C. Niemeyer and James W. Truran. Cambridge University Press, 2000. Available at **astro-ph/9904049**

A Cyclic Model of the Universe. Paul J. Steinhardt and Neil Turok in *Science,* Vol. 296, No. 5572, pages 1436–1439; 2002.

EXPANSION

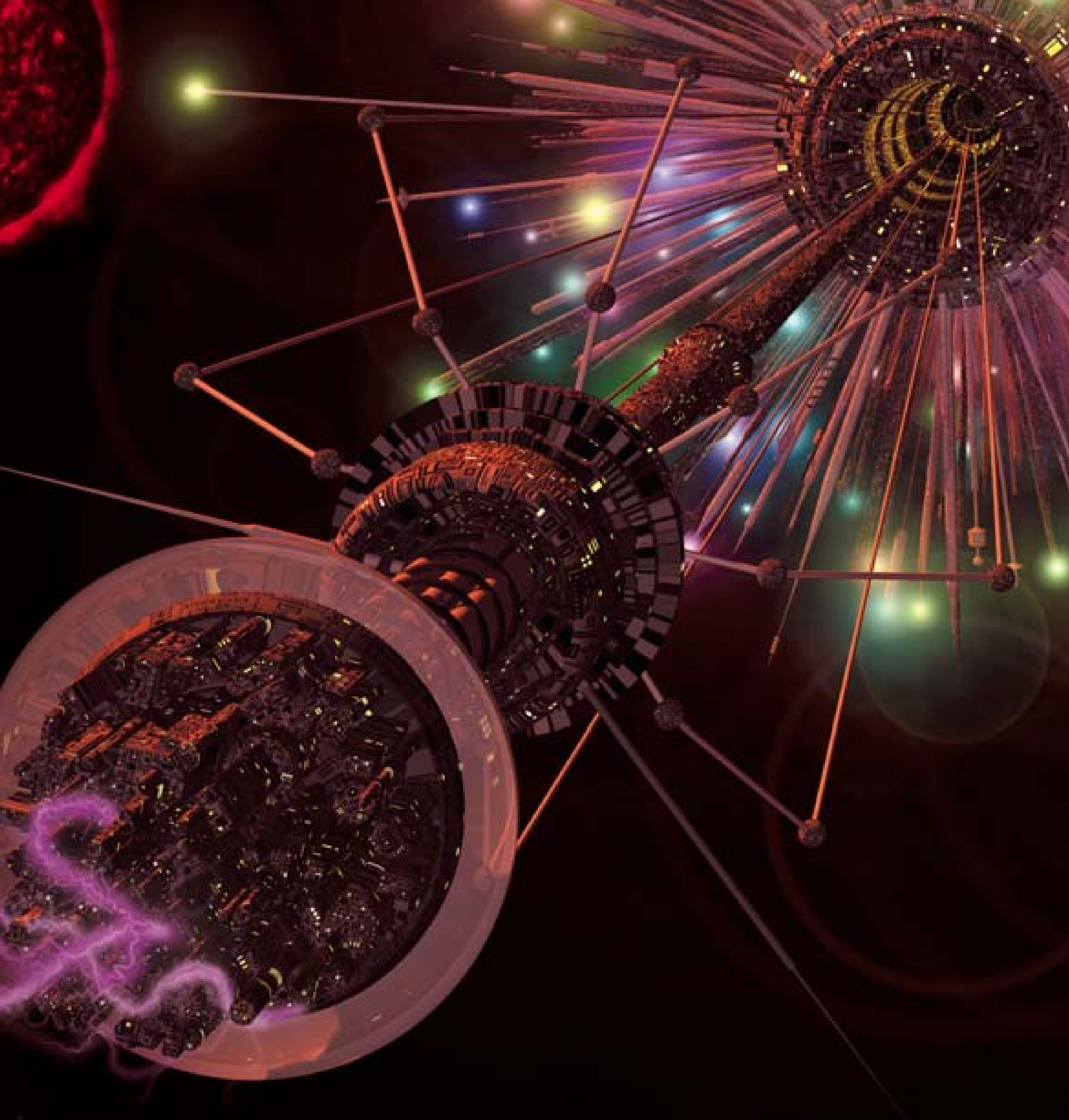
the fate of life in the universe

BILLIONS OF YEARS AGO THE UNIVERSE WAS TOO HOT FOR LIFE TO EXIST. COUNTLESS AEONS FROM NOW, IT WILL BECOME SO COLD AND EMPTY THAT LIFE, NO MATTER HOW INGENIOUS, WILL PERISH

BY LAWRENCE M. KRAUSS AND GLENN D. STARKMAN

Eternal life is a core belief of many of the world's religions. Usually it is extolled as a spiritual Valhalla, an existence without pain, death, worry or evil, a world removed from our physical reality. But there is another sort of eternal life that we hope for, one in the temporal realm. In the conclusion to *On the Origin of Species,* Charles Darwin wrote: "As all the living forms of life are the lineal descendants of those which lived before the Cambrian epoch, we may feel certain that the ordinary succession by generation has never once been broken.... Hence we may look with some confidence to a secure future of great length." The sun will eventually exhaust its hydrogen fuel, and life as we know it on our home planet will eventually end, but the human race is resilient. Our progeny will seek new homes, spreading into every corner of the universe just as organisms have colonized every possible niche of the earth. Death and evil will take their toll,

SLIM FILMS; SOURCES: FRED C. ADAMS AND GREGORY LAUGHLIN *University of Michigan at Ann Arbor;* LAWRENCE M. KRAUSS AND GLENN D. STARKMAN; TIME LINE BY MARK CLEMENS



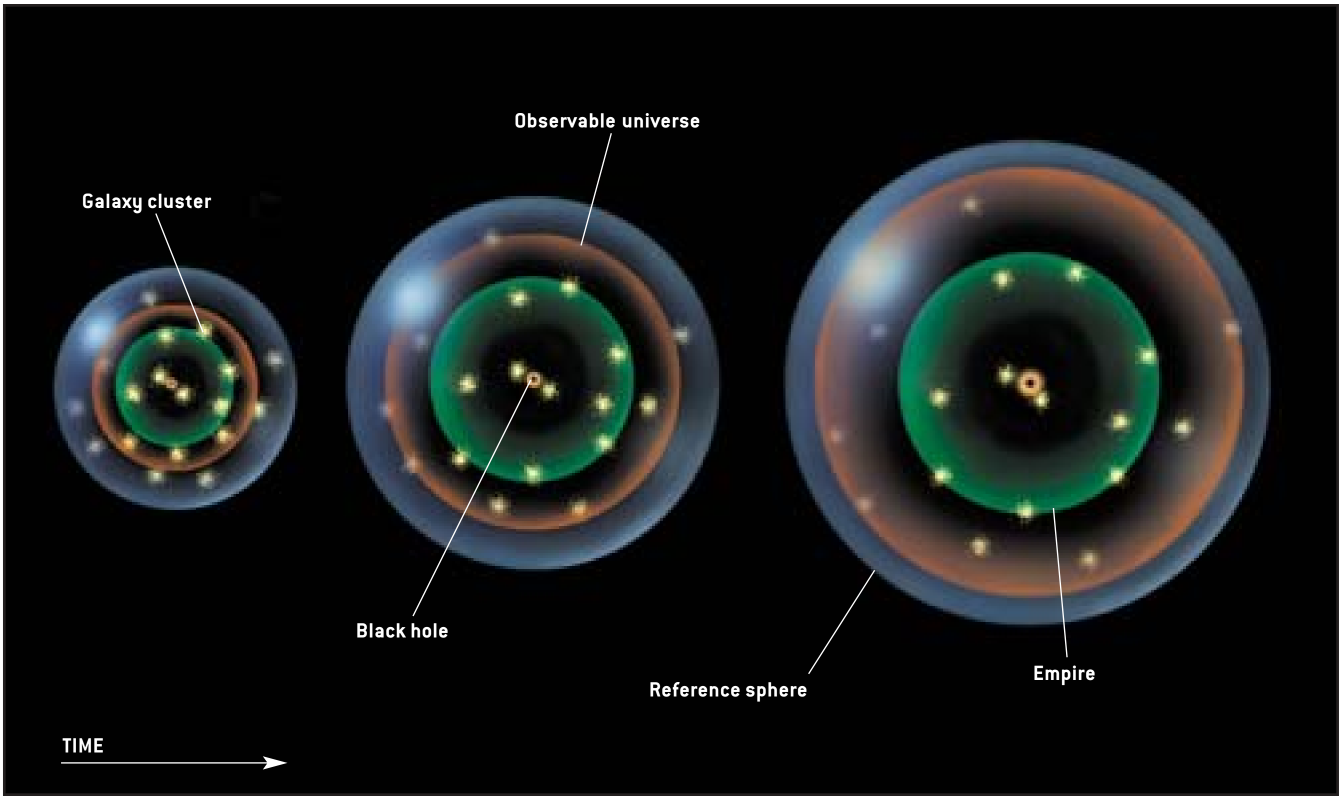
MILESTONES ON THE ROAD TO ETERNITY range from the big bang through the birth and death of stars (*time line below*). As the last stars wane, intelligent beings will need to find new sources of energy, such as cosmic strings (*illustration above*). Unfortunately, natural processes—such as outbreaks of black holes—will erode these linear concentrations of energy, eventually forcing life-forms to seek sustenance elsewhere, if they can find it. Because the governing processes of the universe act on widely varying timescales, the time line is best given a logarithmic scale. If the universe is now expanding at an accelerating rate, additional effects (*shown on time line in blue*) will make life even more miserable.

Space and time disentangle

10^{-51} year since big bang

Cosmic inflation

10^{-44}



ENERGY COLLECTION STRATEGY devised by physicist Steven Frautschi of Caltech illustrates how difficult it will be to survive in the far future, 10^{100} or so years from now. In many cosmological scenarios, resources multiply as the universe—and any arbitrary reference sphere within it (*blue sphere*)—expands and an increasing fraction of it becomes observable (*red sphere*). A civilization could use a black hole to convert matter—plundered from its empire (*green sphere*)—into energy. But as the empire grows, the cost of capturing new territory increases; the conquest can barely keep pace with the dilution of matter. In fact, matter will become so diluted that the civilization will not be able to safely build a black hole large enough to collect it.

DON DIXON AND GEORGE MUSSER

pain and worry may never go away, but somewhere we expect that some of our children will carry on.

Or maybe not. Remarkably, even though scientists fully understand neither the physical basis of life nor the unfolding of the universe, they can make educated guesses about the destiny of living things. Cosmological observations now suggest that the universe will continue to expand forever—rather than, as scientists once thought, expanding to a maximum size and then shrinking. Therefore, we are not doomed to perish in a fiery "big crunch" in which any vestige of our current or future civilization would be erased. At first glance, eternal expansion is cause for optimism. What could stop a sufficiently intelligent civilization from exploiting the endless resources to survive indefinitely?

The Deserts of Vast Eternity

YET LIFE THRIVES ON energy and information, and very general scientific arguments hint that only a finite amount of energy and a finite amount of information can be amassed in even an infinite period. For life to persist, it would have to make do with dwindling resources and limited knowledge. We have concluded that no meaningful form of consciousness could exist forever under these conditions.

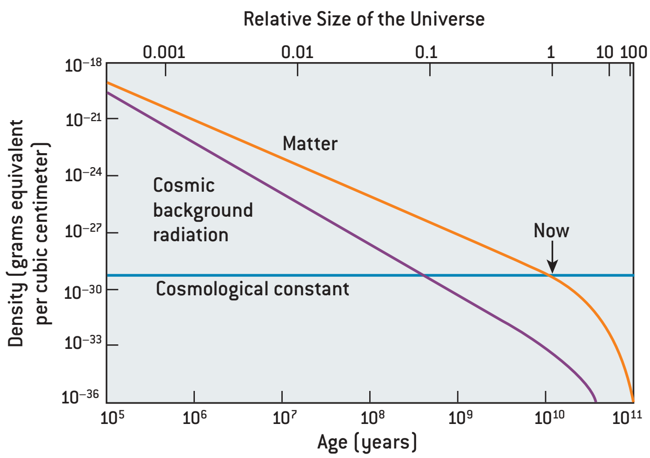
Over the past century, scientific eschatology has swung between optimism and pessimism. Not long after Darwin's confident prediction, Victorian-era scientists began to fret about the "heat death," in which the whole cosmos would come to a common temperature and thereafter be incapable of change. The discovery of the expansion of the universe in the 1920s allayed this concern, because expansion prevents the universe from reaching such an equilibrium. But few cosmologists thought through the other implications for life in an ever expanding universe, until a classic paper in 1979 by physicist Freeman Dyson of the Institute for Advanced Study in Princeton, N.J., itself motivated by earlier work by Jamal Islam, now at the University of Chittagong in Bangladesh. Since Dyson's paper, physicists and astronomers have periodically reexamined the topic [see "The Future of the Universe," by Duane A. Dicus, John R. Letaw, Doris C. Teplitz and Vigdor L. Teplitz; SCIENTIFIC AMERICAN, March 1983]. In 1998, spurred on by new observations that suggest a drastically different long-term future for the universe than that previously envisaged, we began to take another look.

Over the past 12 billion years or so, the universe has passed through many stages. At the earliest times for which scientists now have empirical information, it was incredibly hot and

Electromagnetism emerges
10^{-18} year since big bang

Atomic nuclei created
10^{-5}

Updated from the November 1999 issue



DILUTION of the cosmos by the expansion of space affects different forms of energy in different ways. Ordinary matter (*orange*) thins out in direct proportion to volume, whereas the cosmic background radiation (*purple*) weakens even faster as it is stretched from light into microwaves and beyond. The energy density represented by a cosmological constant (*blue*) does not change, at least according to present theories.

dense. Gradually, it expanded and cooled. For hundreds of thousands of years, radiation ruled; the famous cosmic microwave background radiation is thought to be a vestige of this era. Then matter started to dominate, and progressively larger astronomical structures condensed out. Now, if recent cosmological observations are correct, the expansion of the universe is beginning to accelerate—a sign that a strange new type of energy, perhaps springing from space itself, may be taking over.

Life as we know it depends on stars. But stars inevitably die, and their birth rate has declined dramatically since an initial burst about 10 billion years ago. About 100 trillion years from now, the last conventionally formed star will wink out, and a new era will commence. Processes currently too slow to be noticed will become important: the dispersal of planetary systems by stellar close encounters, the possible decay of ordinary and exotic matter, the slow evaporation of black holes.

Assuming that intelligent life can adapt to the changing circumstances, what fundamental limits does it face? In an eternal universe, potentially of infinite volume, one might hope that a sufficiently advanced civilization could collect an infinite amount of matter, energy and information. Surprisingly, this is not true. Even after an eternity of hard and well-planned labor, living beings could accumulate only a finite number of particles, a finite quantity of energy and a finite number of bits of information. What makes this failure all the more frustrating is that the number of available particles, ergs and bits may grow without bound. The problem is not necessarily the lack of resources but rather the difficulty in collecting them.

The culprit is the very thing that allows us to contemplate an eternal tenure: the expansion of the universe. As the cosmos grows in size, the average density of ordinary sources of energy declines. Doubling the radius of the universe decreases the density of atoms eightfold. For light waves, the decline is even more precipitous. Their energy density drops by a factor of 16 because the expansion stretches them and thereby saps their energy [*see illustration at left*].

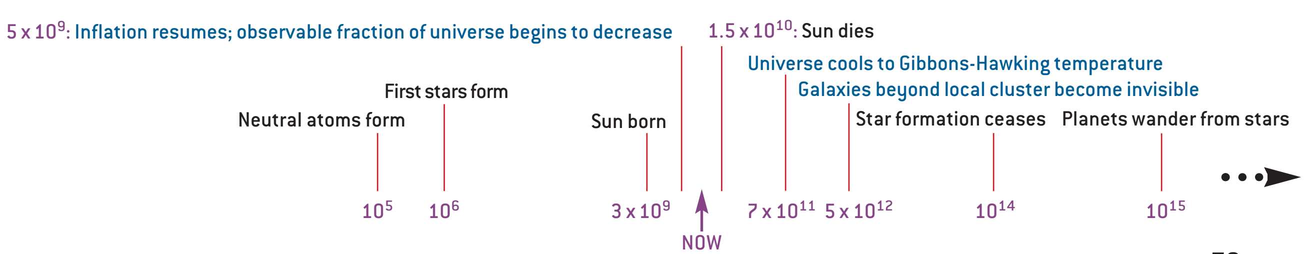
As a result of this dilution, resources become ever more time-consuming to collect. Intelligent beings have two distinct strategies: let the material come to them or try to chase it down. For the former, the best approach in the long run is to let gravity do the work. Of all the forces of nature, just gravity and electromagnetism can draw things in from arbitrarily far away. But the latter gets screened out: oppositely charged particles balance one another, so that the typical object is neutral and hence immune to long-range electrical and magnetic forces. Gravity, on the other hand, cannot be screened out, because particles of matter and radiation only attract gravitationally; they do not repel.

Surrender to the Void

EVEN GRAVITY, however, must contend with the expansion of the universe, which pulls objects apart and thereby weakens their mutual attraction. In all but one scenario, gravity eventually becomes unable to pull together larger quantities of material. Indeed, our universe may have already reached this point; clusters of galaxies may be the largest bodies that gravity will ever be able to bind together [see "The Evolution of Galaxy Clusters," by J. Patrick Henry, Ulrich G. Briel and Hans Böhringer; Scientific American, December 1998]. The lone exception occurs if the universe is poised between expansion and contraction, in which case gravity continues indefinitely to assemble increasingly greater amounts of matter. But that scenario is now thought to contradict observations, and in any

THE AUTHORS

LAWRENCE M. KRAUSS and *GLENN D. STARKMAN* consider their ruminations on the future of life to be a natural extension of their interest in the fundamental workings of the universe. Krauss's books on the predictions of science fiction, *The Physics of Star Trek* and *Beyond Star Trek,* have a similar motivation. The chair of the physics department at Case Western Reserve University, Krauss was among the first cosmologists to argue forcefully that the universe is dominated by a cosmological constant—a view now widely shared. Starkman, also a professor at Case Western, is perhaps best known for his work on the topology of the universe. Both authors are frustrated optimists. They have sought ways that life could persist forever, to no avail. Nevertheless, they maintain the hope that the Cleveland Indians will win the World Series in the ample time that remains.



LAURIE GRACE AND GEORGE MUSSER

event it poses its own difficulty: after 10^{33} years or so, the accessible matter will become so concentrated that most of it will collapse into black holes, sweeping up any life-forms. Being inside a black hole is not a happy condition. On the earth, all roads may lead to Rome, but inside a black hole, all roads lead in a finite amount of time to the center of the hole, where death and dismemberment are certain.

Sadly, the strategy of actively seeking resources fares no better than the passive approach does. The expansion of the universe drains away kinetic energy, so prospectors would have to squander their booty to maintain their speed. Even in the most optimistic scenario—in which the energy is traveling toward the scavenger at the speed of light and is collected without loss—a civilization could garner limitless energy only in or near a black hole. The latter possibility was explored by Steven Frautschi of the California Institute of Technology in 1982. He concluded that the energy available from the holes would dwindle more quickly than the costs of scavenging [*see illustration on page 52*]. We recently reexamined this possibility and found that the predicament is even worse than Frautschi thought. The size of a black hole required to sweep up energy forever exceeds the extent of the visible universe.

The cosmic dilution of energy is truly dire if the universe is expanding at an accelerating rate. All distant objects that are currently in view will eventually move away from us faster than the speed of light and, in doing so, disappear from view. The total resources at our disposal are therefore limited by what we can see today, at most [*see box at right*].

Not all forms of energy are equally subject to the dilution. The universe might, for example, be filled with a network of cosmic strings—infinitely long, thin concentrations of energy that could have developed as the early universe cooled unevenly. The energy per unit length of a cosmic string remains unchanged despite cosmic expansion [see "Cosmic Strings," by Alexander Vilenkin; SCIENTIFIC AMERICAN, December 1987]. Intelligent beings might try to cut one, congregate around the loose ends and begin consuming it. If the string network is infinite, they might hope to satisfy their appetite forever. The problem with this strategy is that whatever life-forms can do, natural processes can also do. If a civilization can figure out a way to cut cosmic strings, then the string network will fall apart of its own accord. For example, black holes may spontaneously appear on the strings and devour them. Therefore, the beings could swallow only a finite amount of string before running into another loose end. The entire string network would eventually disappear, leaving the civilization destitute.

What about mining the quantum vacuum? After all, the cosmic acceleration may be driven by the so-called cosmological constant, a form of energy that does not dilute as the universe expands [see "Cosmological Antigravity," on page 30]. If this is so, empty space is filled with a bizarre type of radiation, called Gibbons-Hawking or de Sitter radiation. Alas, it is impossible to extract energy from this radiation for useful work. If the quantum vacuum yielded up energy, it would drop into a lower energy state, yet the vacuum is already the lowest energy state there is.

No matter how clever we try to be and how cooperative the universe is, we will someday have to confront the finiteness of the resources at our disposal. Even so, are there ways to cope forever?

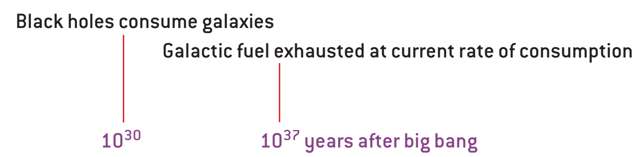
The Worst of All Possible Universes

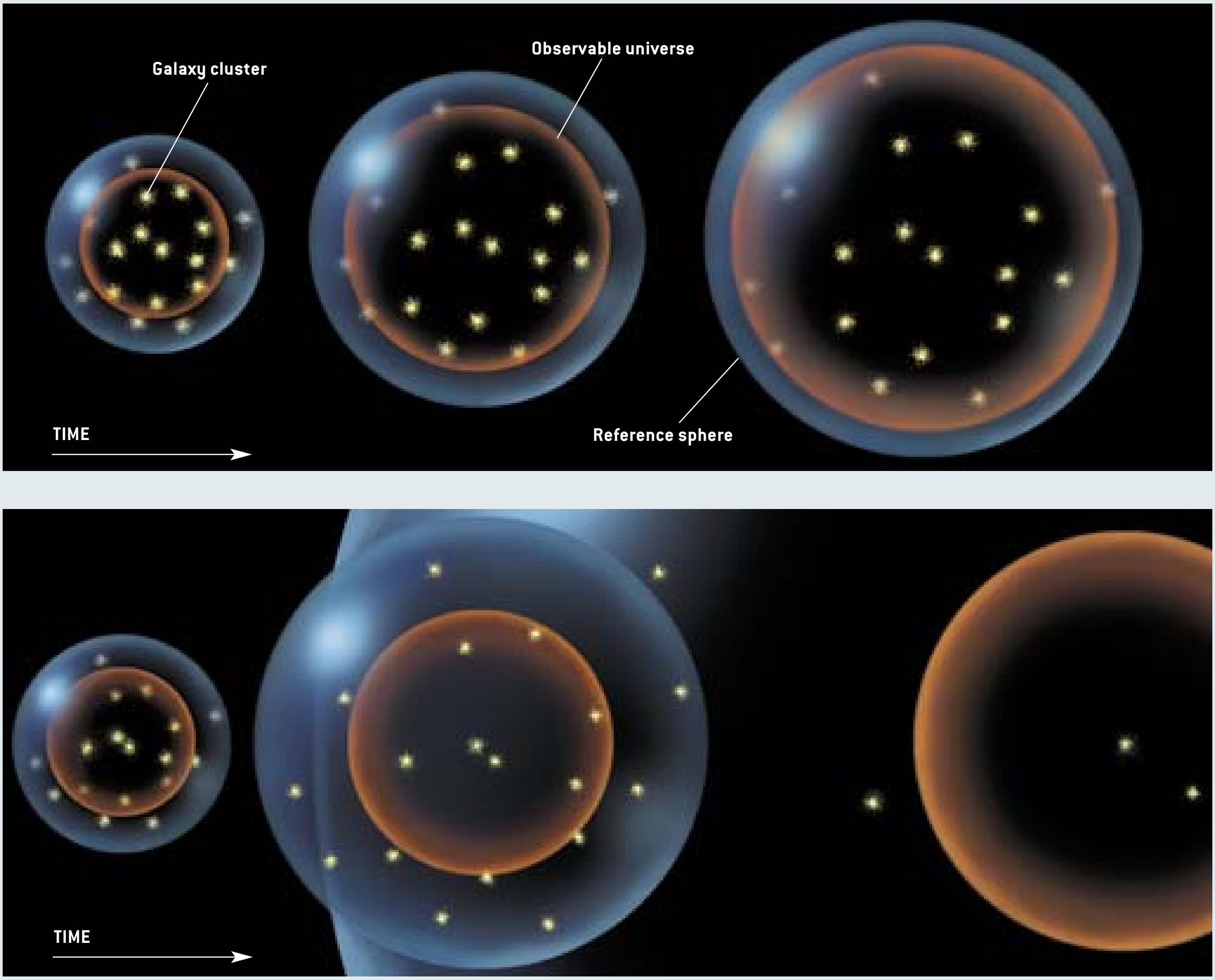
AMONG ALL THE SCENARIOS for an eternally expanding universe, the one dominated by the so-called cosmological constant is the bleakest. Not only is it unambiguous that life cannot survive eternally in such a universe, but the quality of life will quickly deteriorate as well. So if recent observations that the expansion is accelerating are borne out [see "Surveying Spacetime with Supernovae," on page 22], humankind could face a grim future.

Cosmic expansion carries objects away from one another unless they are bound together by gravity or another force. In our case, the Milky Way is part of a larger cluster of galaxies. About 10 million light-years across, this cluster remains a cohesive whole, whereas galaxies beyond it are whisked away as intergalactic space expands. The relative velocity of these distant galaxies is proportional to their distance. Beyond a certain distance called the horizon, the velocity exceeds the speed of light (which is allowed in the general theory of relativity because the velocity is imparted by the expansion of space itself). We can see no farther.

If the universe has a cosmological constant with a positive value, as the observations suggest, the expansion is accelerating: galaxies are beginning to move apart ever more rapidly. Their velocity is still proportional to their distance, but the constant of proportionality remains constant rather than decreasing with time, as it does if the universe decelerates. Consequently, galaxies that are now beyond our horizon will forever remain out of sight. Even the galaxies we can currently see—except for those in the local cluster—will eventually attain the speed of light and vanish from view. The acceleration, which resembles inflation in the very early universe, began when the cosmos was about half its present age.

The disappearance of distant galaxies will be gradual. Their light will stretch out until it becomes undetectable. Over time, the amount of matter we can see will decrease, and the number of worlds our starships can reach will diminish. Within two trillion years, well before the last stars in the universe die, all objects outside our own cluster of galaxies will no longer be observable or accessible. There will be no new worlds to conquer, literally. We will truly be alone in the universe. *—L.M.K. and G.D.S.*





EXPANDING UNIVERSE looks dramatically different depending on whether the growth is decelerating (*upper sequence*) or accelerating (*lower sequence*). In both these cases, the universe is infinite, but any patch of space—demarcated by a reference sphere that represents the distance to particular galaxies—enlarges (*blue sphere*). Humans can see only a limited volume of the universe around them, which grows steadily as light signals have time to propagate (*red sphere*). If expansion is decelerating, we can see an increasing fraction of the cosmos. More and more galaxies fill the sky. But if expansion is accelerating, we can see a decreasing fraction of the cosmos. Space seems to empty out.

DON DIXON AND GEORGE MUSSER

The obvious strategy is to learn to make do with less, a scheme first discussed quantitatively by Dyson. In order to reduce energy consumption and keep it low despite exertion, we would eventually have to reduce our body temperature. One might speculate about genetically engineered humans who function at somewhat lower temperatures than 310 kelvins (98.6 degrees Fahrenheit). Yet the human body temperature cannot be reduced arbitrarily; the freezing point of blood is a firm lower limit. Ultimately, we will need to abandon our bodies entirely.

Though futuristic, the idea of shedding our bodies presents no fundamental difficulties. It presumes only that consciousness is not tied to a particular set of organic molecules but rather can be embodied in a multitude of different forms, from cyborgs to sentient interstellar clouds [see "Will Robots Inherit the Earth?" by Marvin Minsky; Scientific American, October 1994]. Most modern philosophers and cognitive scientists regard conscious thought as a process that a computer could perform. The details need not concern us here (which is convenient, as we are not competent to discuss them). We still have many billions of years to design new physical incarnations to which we will someday transfer our conscious selves. These

Quantum tunneling liquefies matter

10^{65}

new "bodies" will be required to operate at cooler temperatures and at lower metabolic rates—that is, at lower rates of energy consumption.

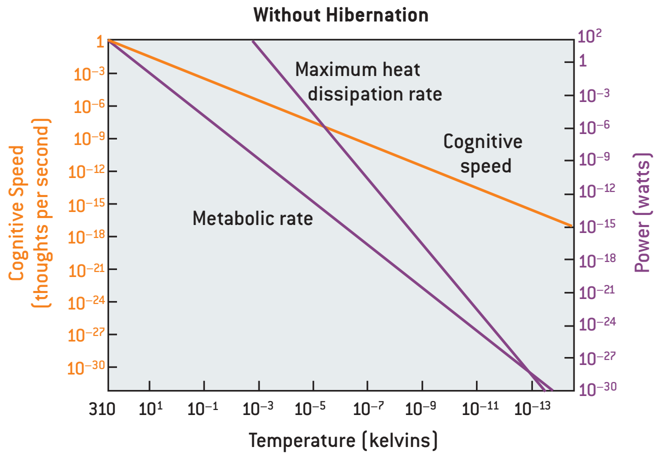
Dyson showed that if organisms could slow their metabolism as the universe cooled, they could arrange to consume a finite total amount of energy over all of eternity. Although the lower temperatures would also slow consciousness—the number of thoughts per second—the rate would remain large enough for the total number of thoughts, in principle, to be unlimited. In short, intelligent beings could survive forever, not just in absolute time but in subjective time as well. As long as organisms were guaranteed to have an infinite number of thoughts, they would not mind a languid pace of life. When billions of years stretch out before you, what's the rush?

At first glance, this might look like a case of something for nothing. But the mathematics of infinity can defy intuition. For an organism to maintain the same degree of complexity, Dyson argued, its rate of information processing must be directly proportional to body temperature, whereas the rate of energy consumption is proportional to the square of the temperature (the additional factor of temperature comes from basic thermodynamics). Therefore, the power requirements slacken faster than cognitive alacrity does [*see illustration at right*]. At 310 kelvins, the human body expends approximately 100 watts. At 155 kelvins, an equivalently complex organism could think at half the speed but consume a quarter of the power. The trade-off is acceptable because physical processes in the environment slow down at a similar rate.

To Sleep, to Die

UNFORTUNATELY, there is a catch. Most of the power is dissipated as heat, which must escape—usually by radiating away—if the object is not to heat up. Human skin, for example, glows in infrared light. At very low temperatures, the most efficient radiator would be a dilute gas of electrons. Yet the efficiency of even this optimal radiator declines as the cube of the temperature, faster than the decrease in the metabolic rate. A point would come when organisms could not lower their temperature further. They would be forced instead to reduce their complexity—to dumb down. Before long, they could no longer be regarded as intelligent.

To the timid, this might seem like the end. But to compensate for the inefficiency of radiators, Dyson boldly devised a strategy of hibernation. Organisms would spend only a small fraction of their time awake. While sleeping, their metabolic rates would drop, but—crucially—they would continue to dissipate heat. In this way, they could achieve an ever lower average body temperature [*see illustration on opposite page*]. In fact, by spending an increasing fraction of their time asleep,



ETERNAL LIFE ON FINITE ENERGY? If a new form of life could lower its body temperature below the human value of 310 kelvins (98.6 degrees Fahrenheit), it would consume less power, albeit at the cost of thinking more sluggishly (*left graph*). Because metabolism would decline faster than cognition, the life-form could arrange to have an infinite number of thoughts on limited resources. One caveat is that its ability to dissipate waste heat would also decline, preventing it from cooling below about

they could consume a finite amount of energy yet exist forever and have an infinite number of thoughts. Dyson concluded that eternal life is indeed possible.

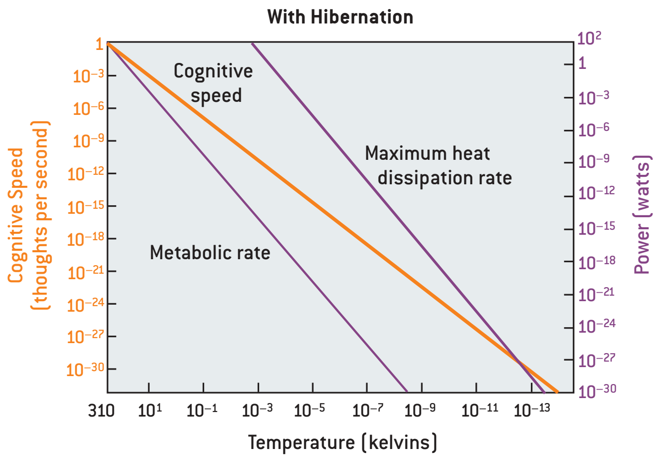
Since his original paper, several difficulties with his plan have emerged. For one, Dyson assumed that the average temperature of deep space—currently 2.7 kelvins, as set by the cosmic microwave background radiation—would always decrease as the cosmos expands, so that organisms could continue to decrease their temperature forever. But if the universe has a cosmological constant, the temperature has an absolute floor fixed by the Gibbons-Hawking radiation. For current estimates of the value of the cosmological constant, this radiation has an effective temperature of about 10^{-29} kelvin. As was noted independently by cosmologists J. Richard Gott II, John Barrow, Frank Tipler and us, once organisms had cooled to this level, they could not continue to lower their temperature in order to conserve energy.

The second difficulty is the need for alarm clocks to wake the organisms periodically. These clocks would have to operate reliably for longer and longer times on less and less energy. Quantum mechanics suggests that this is impossible. Consider, for example, an alarm clock that consists of two small balls that are taken far apart and then aimed at each other and released. When they collide, they ring a bell. To lengthen the time between alarms, organisms would release the balls at a slower speed. But eventually the clock will run up against constraints from Heisenberg's uncertainty principle, which prevents the speed and position of the balls from both being specified to arbitrary precision. If one or the other is sufficient-

LAURIE GRACE AND GEORGE MUSSER

Electrons and positrons bind into new form of matter — 10^{85} years after big bang

Galactic black holes evaporate — 10^{98}



10^{-13} kelvin. Hibernation (*right graph*) might eliminate the problem of heat disposal. As the life-form cooled, it would spend an increasing fraction of its time dormant, further reducing its average metabolic rate and cognitive speed. In this way, the power consumption could always remain lower than the maximum rate of heat dissipation, while still allowing for an infinite number of thoughts. But such a scheme might run afoul of other problems, such as quantum limits.

ly inaccurate, the alarm clock will fail, and hibernation will turn into eternal rest.

One might imagine other alarm clocks that could forever remain above the quantum limit and might even be integrated into the organism itself. Nevertheless, no one has yet come up with a specific mechanism that could reliably wake an organism while consuming finite energy.

The Eternal Recurrence of the Same

THE THIRD AND MOST general doubt about the long-term viability of intelligent life involves fundamental limitations on computation. Computer scientists once thought it was impossible to compute without expending a certain minimum amount of energy per operation, an amount that is directly proportional to the temperature of the computer. Then, in the early 1980s, researchers realized that certain physical processes, such as quantum effects or the random Brownian motion of a particle in a fluid, could serve as the basis for a lossless computer [see "The Fundamental Physical Limits of Computation," by Charles H. Bennett and Rolf Landauer; Scientific American, July 1985].

Such computers could operate with an arbitrarily small amount of energy. To use less, they simply slow down—a trade-off that eternal organisms may be able to make. There are only two conditions. First, they must remain in thermal equilibrium with their environment. Second, they must never discard information. If they did, the computation would become irreversible, and thermodynamically an irreversible process must dissipate energy.

Unhappily, those conditions become insurmountable in an expanding universe. As cosmic expansion dilutes and stretches the wavelength of light, organisms become unable to emit or absorb the radiation they would need to establish thermal equilibrium with their surroundings. And with a finite amount of material at their disposal, and hence a finite memory, they would eventually have to forget an old thought in order to have a new one. What kind of perpetual existence could such organisms have, even in principle? They could collect just a finite number of particles and a finite amount of information. Those particles and bits could be configured in just a finite number of ways. Because thoughts are the reorganization of information, finite information implies a finite number of thoughts. All organisms would ever do is relive the past, having the same thoughts over and over again. Eternity would become a prison, rather than an endlessly receding horizon of creativity and exploration. It might be nirvana, but would it be living?

It is only fair to point out that Dyson has not given up. In his correspondence with us, he has suggested that life can avoid the quantum constraints on energy and information by, for example, growing in size or using different types of memory. As he intriguingly puts it, the question is whether life is "analog" or "digital"—that is, whether continuum physics or quantum physics sets its limits. We believe that over the long haul, life is digital.

Is there any other hope for eternal life? Quantum mechanics, which we argue puts such unbending limits on life, might come to its rescue in another guise. For instance, if the quantum mechanics of gravity allows the existence of stable wormholes, life-forms might circumvent the barriers erected by the speed of light, visit parts of the universe that are otherwise inaccessible, and collect infinite amounts of energy and information. Or perhaps they could construct "baby" universes [see "The Self-Reproducing Inflationary Universe," by Andrei Linde; Scientific American, November 1994] and send themselves—or at least a set of instructions that could be used to reconstitute themselves—through to the baby universe. In that way, life could carry on.

The ultimate limits on life will in any case become significant only on timescales that are truly cosmic. Still, to some it may seem disturbing that life, certainly in its physical incarnation, must come to an end. But to us, it is remarkable that even with our limited knowledge, we can draw conclusions about such grand issues. Perhaps being cognizant of our fascinating universe and our destiny within it is a greater gift than being able to inhabit it forever. SA

MORE TO EXPLORE

Time without End: Physics and Biology in an Open Universe. Freeman J. Dyson in *Reviews of Modern Physics,* Vol. 51, No. 3, pages 447–460; July 1979.

The Anthropic Cosmological Principle. John D. Barrow and Frank J. Tipler. Oxford University Press, 1988.

The Last Three Minutes: Conjectures about the Ultimate Fate of the Universe. Paul C. W. Davies. HarperCollins, 1997.

The Five Ages of the Universe: Inside the Physics of Eternity. Fred Adams and Greg Laughlin. Free Press, 1999.

Quintessence: The Mystery of the Missing Mass. Lawrence M. Krauss. Basic Books, 1999.

Life, the Universe, and Nothing: Life and Death in an Ever-Expanding Universe. Lawrence M. Krauss and Glenn D. Starkman in *Astrophysical Journal,* Vol. 531, pages 22–30; 2000. Available at xxx.lanl.gov/abs/astro-ph/9902189

STRUCTURE

Is Space Finite?

Conventional wisdom says the universe is infinite. But it could be finite, merely giving the illusion of infinity

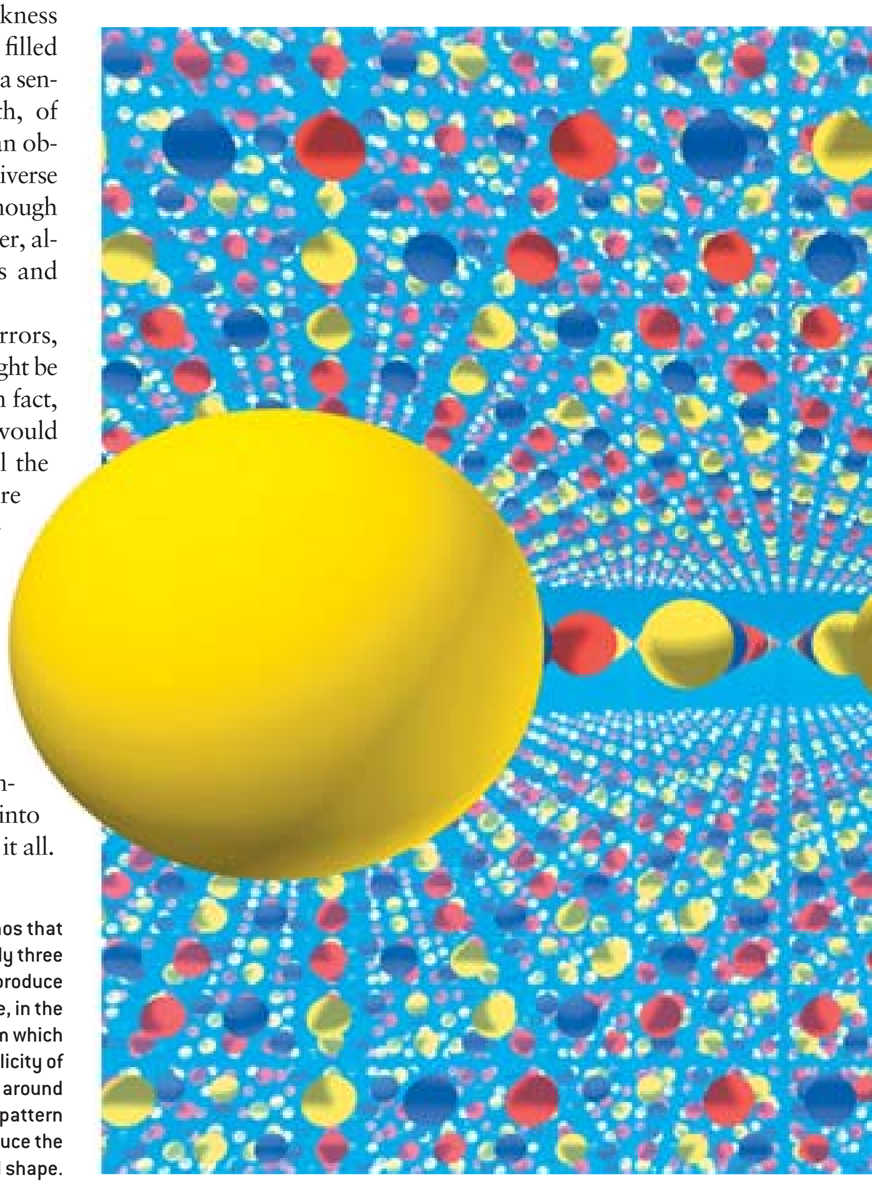
BY JEAN-PIERRE LUMINET, GLENN D. STARKMAN AND JEFFREY R. WEEKS

Illustrations by Bryan Christie Design

Looking up at the sky on a clear night, we feel we can see forever. There seems to be no end to the stars and galaxies; even the darkness in between them is filled with light if only we stare through a sensitive enough telescope. In truth, of course, the volume of space we can observe is limited by the age of the universe and the speed of light. But given enough time, could we not peer ever farther, always encountering new galaxies and phenomena?

Maybe not. Like a hall of mirrors, the apparently endless universe might be deluding us. The cosmos could, in fact, be finite. The illusion of infinity would come about as light wrapped all the way around space, perhaps more than once—creating multiple images of each galaxy. Our own Milky Way galaxy would be no exception; bizarrely, the skies might even contain facsimiles of the earth at some earlier era. As time marched on, astronomers could watch the galaxies develop and look for new mirages. But eventually no new space would enter into their view. They would have seen it all.

The question of a finite or infinite universe is one of the oldest in philosophy. A common misconception is that it has already been settled in favor of the latter. The reasoning, often repeated in textbooks, draws an unwarranted conclusion from

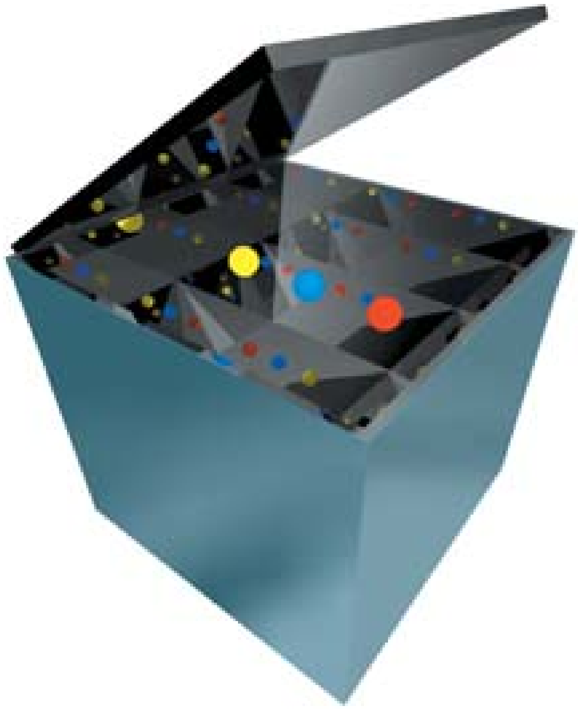


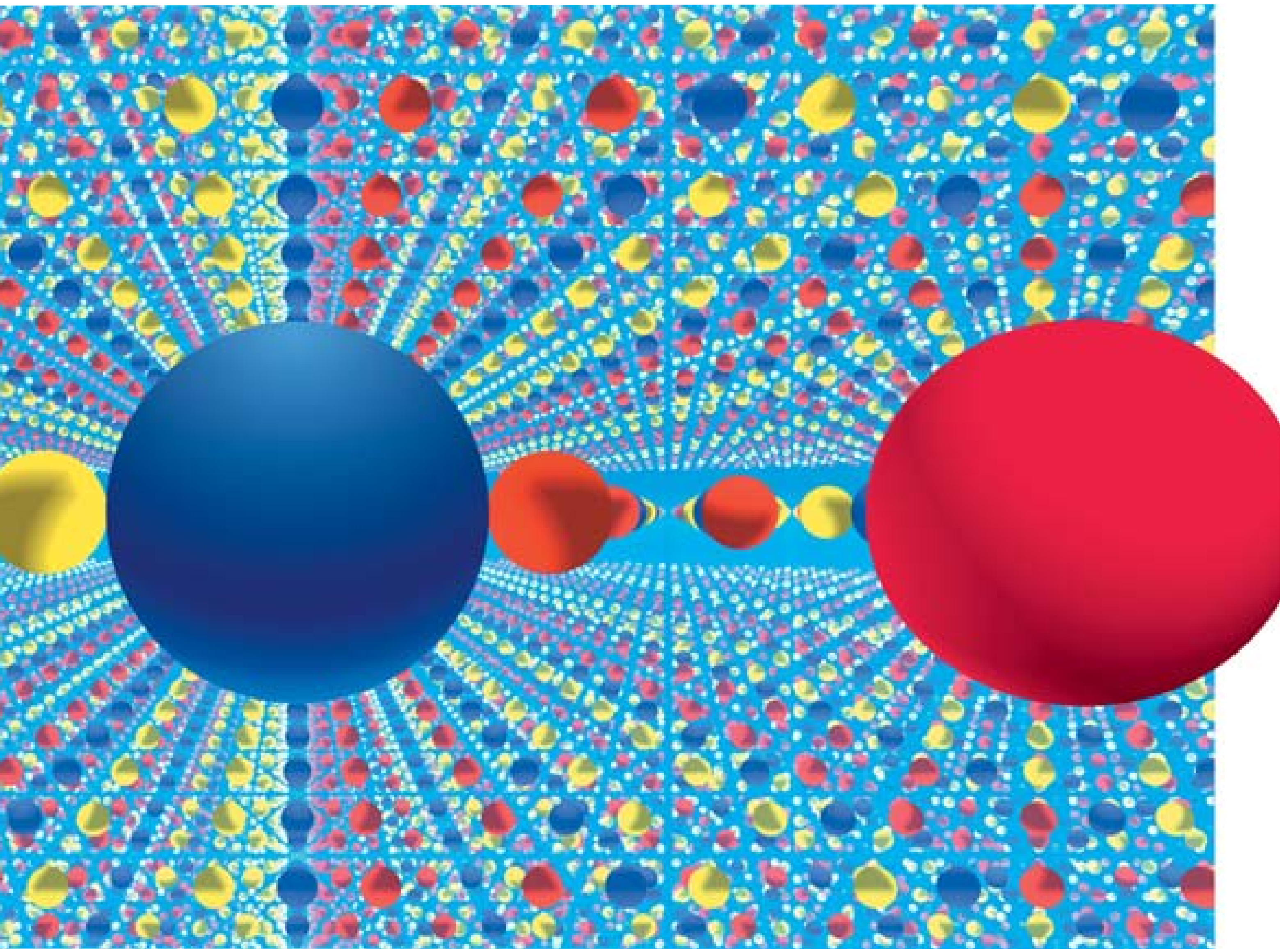
"INFINITY BOX" evokes a finite cosmos that looks endless. The box contains only three balls, yet the mirrors that line its walls produce an infinite number of images. Of course, in the real universe there is no boundary from which light can reflect. Instead a multiplicity of images could arise as light rays wrap around the universe again and again. From the pattern of repeated images, one could deduce the universe's true size and shape.

Updated from the April 1999 issue

Einstein's general theory of relativity. According to relativity, space is a dynamic medium that can curve in one of three ways, depending on the distribution of matter and energy within it. Because we are embedded in space, we cannot see the flexure directly but rather perceive it as gravitational attraction and geometric distortion of images. To determine which of the three geometries our universe has, astronomers have been measuring the density of matter and energy in the cosmos. It now appears to be too low to force space to arch back on itself—a "spherical" geometry. Therefore, space must have either the familiar Euclidean geometry, like that of a plane, or a "hyperbolic" geometry, like that of a saddle [*see illustration on next page*]. At first glance, such a universe stretches on forever.

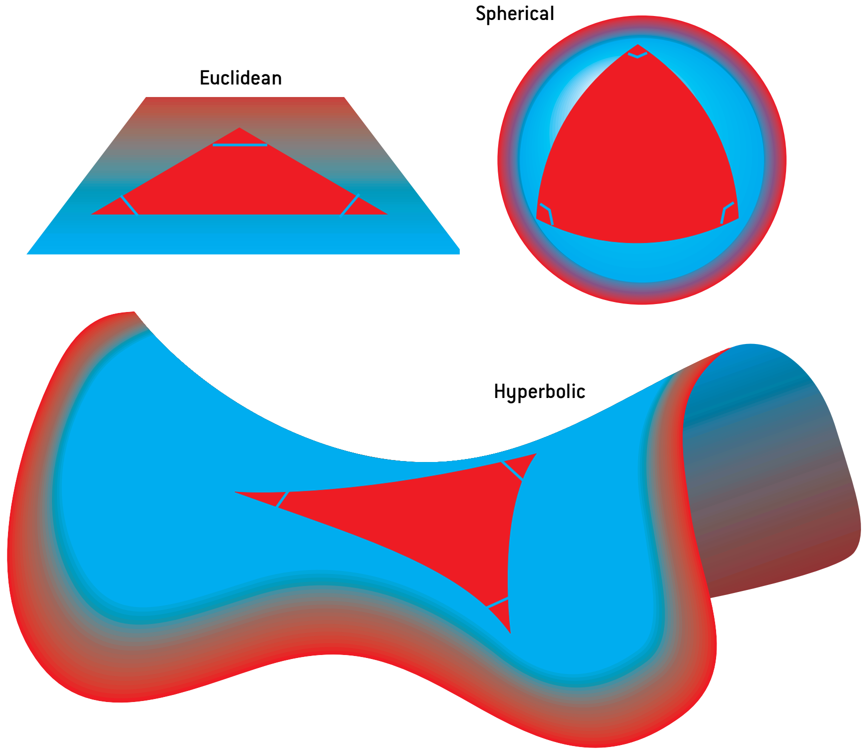
One problem with this conclusion is that the universe could be spherical yet so large that the observable part seems Euclidean, just as a small patch of the earth's surface looks flat. A broader issue, however, is that relativity is a purely local theory. It predicts the curvature of each small volume of space—its





geometry—based on the matter and energy it contains. Neither relativity nor standard cosmological observations say anything about how those volumes fit together to give the universe its overall shape—its topology. The three plausible cosmic geometries are consistent with many different topologies. For example, relativity would describe both a torus (a doughnutlike shape) and a plane with the same equations, even though the torus is finite and the plane is infinite. Determining the topology requires some physical understanding beyond relativity.

The usual assumption is that the universe is, like a plane, "simply connected," which means there is only one direct path for light to travel from a source to an observer. A simply connected Euclidean or hyperbolic universe would indeed be infinite. But the universe might instead be "multiply connected," like a torus, in which case there would be many different paths. An observer would see multiple images of each galaxy and could easily misinterpret them as distinct galaxies in an endless space, much as a visitor to a mirrored room has the illusion of seeing a huge crowd.



LOCAL GEOMETRY of space can be Euclidean, spherical or hyperbolic—the only possibilities consistent with the observed symmetry of the cosmos on large scales. On the Euclidean plane, the angles of a triangle add to exactly 180 degrees; on the spherical surface, they add to more than 180 degrees; and on the hyperbolic surface (or saddle), to less than 180 degrees. Local geometry determines how objects move. But it does not describe how individual volumes connect to give the universe its global shape.

A multiply connected space is no mere mathematical whimsy; it is even preferred by some schemes for unifying the fundamental forces of nature, and it does not contradict any available evidence. Over the past few years, research into cosmic topology has blossomed. New observations may soon reach a definitive answer.

Comfort in the Finite

MANY COSMOLOGISTS EXPECT the universe to be finite. Part of the reason may be simple comfort: the human mind encompasses the finite more readily than the infinite. But there are also two scientific lines of argument that favor finitude. The first involves a thought experiment devised by Isaac Newton and revisited by George Berkeley and Ernst Mach. Grappling with the causes of inertia, Newton imagined two buckets partially filled with water. The first bucket is left still, and the surface of the water is flat. The second bucket is spun rapidly, and the surface of the water is concave. Why?

The naive answer is centrifugal force. But how does the second bucket know it is spinning? In particular, what defines the inertial reference frame relative to which the second bucket spins and the first does not? Berkeley and Mach's answer was that all the matter in the universe collectively provides the reference frame. The first bucket is at rest relative to distant galaxies, so its surface remains flat. The second bucket spins relative to those galaxies, so its surface is concave. If there were no distant galaxies, there would be no reason to prefer one reference frame over the other. The surface in both buckets would have to remain flat, and therefore the water would require no centripetal force to keep it rotating. In short, it would have no inertia. Mach inferred that the amount of inertia a body experiences is proportional to the total amount of matter in the universe. An infinite universe would cause infinite inertia. Nothing could ever move.

In addition to Mach's argument, there is preliminary work in quantum cosmology, which attempts to describe how the universe emerged spontaneously from the void. Some such theories predict that a low-volume universe is more probable than a high-volume one. An infinite universe would have zero probability of coming into existence [see "Quantum Cosmology and the Creation of the Universe," by Jonathan J. Halliwell; SCIENTIFIC AMERICAN, December 1991]. Loosely speaking, its energy would be infinite, and no quantum fluctuation could muster such a sum.

Historically, the idea of a finite universe ran into its own obstacle: the apparent need for an edge. Aristotle argued that the universe is finite on the grounds that a boundary was necessary to fix an absolute reference frame, which was important to his worldview. But his critics wondered what happened at the edge. Every edge has another side. So why not redefine the "universe" (which roughly means "one side") to include that other side? German mathematician Georg F. B. Riemann solved the riddle in the mid-19th century. As a model for the cosmos, he proposed the hypersphere—the three-dimensional surface of a four-dimensional ball, just as an ordinary sphere is the two-dimensional surface of a three-dimensional ball. It was the first example of a space that is finite yet has no problematic boundary.

One might still ask what is outside the universe. But this question supposes that the ultimate physical reality must be a Euclidean space of some dimension. That is, it presumes that if space is a hypersphere, then that hypersphere must sit in a four-dimensional Euclidean space, allowing us to view it from the outside. Nature, however, need not cling to this notion. It would be perfectly acceptable for the universe to be a hypersphere and not be embedded in any higher-dimensional space. Such an object may be difficult to visualize, because we are used to viewing shapes from the outside. But there need not be an "outside."

By the end of the 19th century, mathematicians had discovered a variety of finite spaces without boundaries. German astronomer Karl Schwarzschild brought this work to the attention of his colleagues in 1900. In a postscript to an article in *Vierteljahrschrift der Astronomischen Gesellschaft,* he challenged his readers:

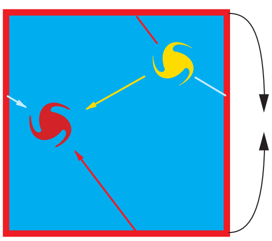
> Imagine that as a result of enormously extended astronomical experience, the entire universe consists of countless identical copies of our Milky Way, that the infinite space can be partitioned into cubes each containing an exactly identical copy of our Milky Way. Would we really cling on to the assumption of infinitely many identical repetitions of the same world?... We would be much happier with the view that these repetitions are illusory, that in reality space has peculiar connection properties so that if we leave any one cube through a side, then we immediately reenter it through the opposite side.

Schwarzschild's example illustrates how one can mentally construct a torus from Euclidean space. In two dimensions, begin with a square and identify opposite sides as the same—as is done in many video games, such as the venerable Asteroids, in which a spaceship going off the right side of the screen reappears on the left side. Apart from the interconnections between sides, the space is as it was before. All the familiar rules of Euclidean geometry hold. At first glance, the space looks infinite to those who live within it, because there is no limit to how far they can see. Without traveling around the universe and reencountering the same objects, the ship could not tell that it is in a torus [*see illustration below*]. In three dimensions, one begins with a cubical block of space and glues together opposite faces to produce a 3-torus.

The Euclidean 2-torus, apart from some sugar glazing, is topologically equivalent to the surface of a doughnut. Unfortunately, the Euclidean torus cannot sit in our three-dimensional Euclidean space. Doughnuts may do so because they have been bent into a spherical geometry around the outside and a hyperbolic geometry around the hole. Without this curvature, doughnuts could not be viewed from the outside.

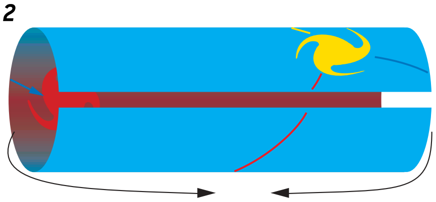
When Albert Einstein published the first relativistic model of the universe in 1917, he chose Riemann's hypersphere as the overall shape. At that time, the topology of space was an active topic of discussion. Russian mathematician Aleksander Friedmann soon generalized Einstein's model to permit an expanding universe and a hyperbolic space. His equations are still routinely used by cosmologists. He emphasized that the equations

1

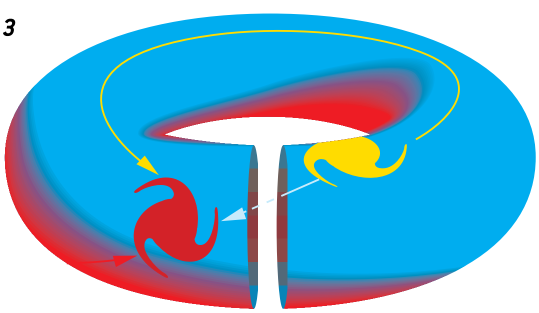


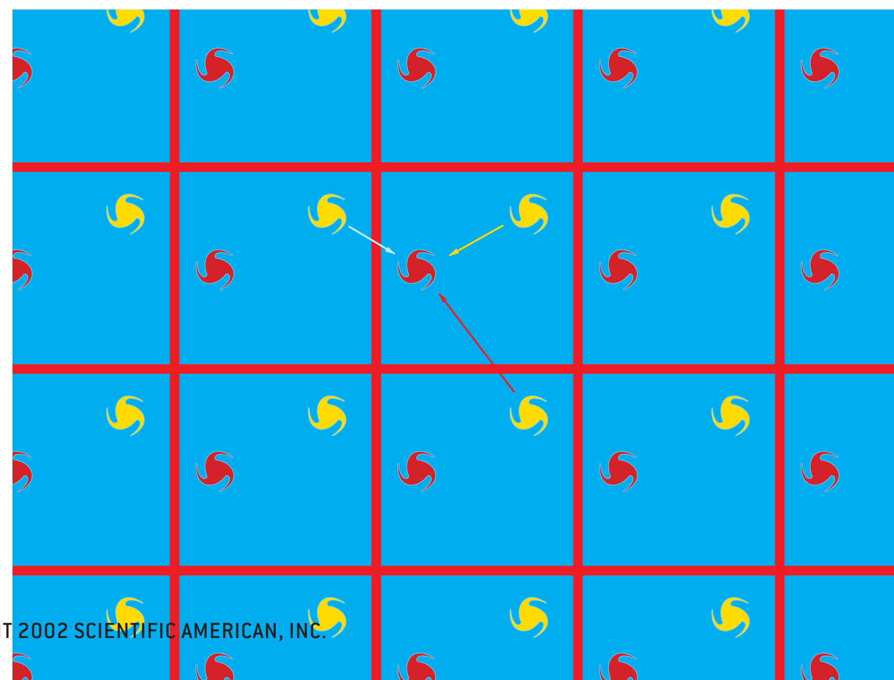
DOUGHNUT SPACE, more properly known as the Euclidean 2-torus, is a flat square whose opposite sides are connected (*1*). Anything crossing one edge reenters from the opposite edge. Although this surface cannot exist within our three-dimensional space, a distorted version can be built by taping together top and bottom (*2*) and scrunching the resulting cylinder into a ring (*3*). For observers in the pictured red galaxy, space seems infinite because their line of sight never ends (*below*). Light from the yellow galaxy can reach them along several different paths, so they see more than one image of it. A Euclidean 3-torus is built from a cube rather than a square.

2



3





of his hyperbolic model applied to finite universes as well as to the standard infinite one—an observation all the more remarkable because, at the time, no examples of finite hyperbolic spaces were known. In fact, almost all topologies require hyperbolic geometries. In two dimensions, a finite Euclidean space must have the topology of either a 2-torus or a Klein bottle; in three dimensions, there are only 10 Euclidean possibilities—namely, the 3-torus and nine simple variations on it, such as gluing together opposite faces with a quarter turn or with a reflection, instead of straight across. By comparison, there are infinitely many possible topologies for a finite hyperbolic three-dimensional universe. Their rich structure is still the subject of intense research. Similarly, there are infinitely many possible topologies for a finite spherical three-dimensional universe.

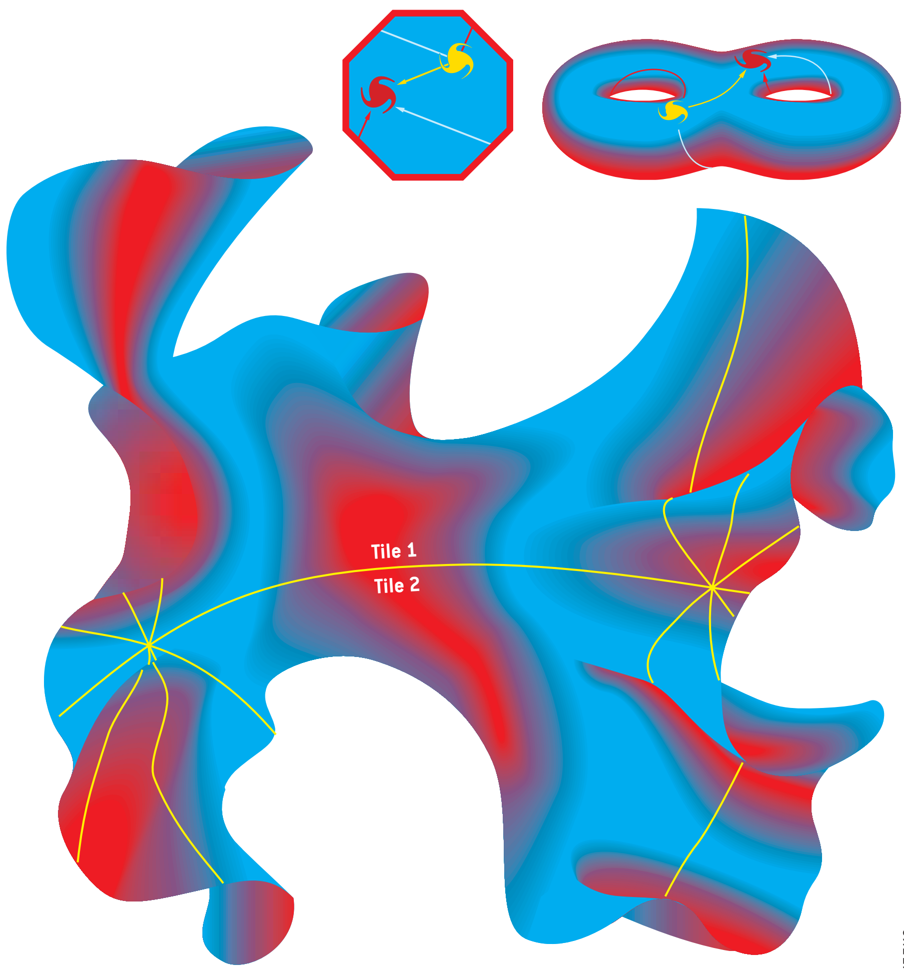
Eightfold

OF ALL THE ISSUES in cosmic topology, perhaps the most difficult to grasp is how a hyperbolic space can be finite. For simplicity, first consider a two-dimensional universe. Mimic the construction of a 2-torus but begin with a hyperbolic surface instead. Cut out a regular octagon and identify opposite pairs of edges, so that anything leaving the octagon across one edge returns at the opposite edge. Alternatively, one could devise an octagonal Asteroids screen [*see illustration at right*]. This is a multiply connected universe, topologically equivalent to a two-holed pretzel. An observer at the center of the octagon sees the nearest images of himself or herself in eight different directions. The illusion is that of an infinite hyperbolic space, even though this universe is really finite. Similar constructions are possible in three dimensions, although they are harder to visualize.

The angles of the octagon merit careful consideration. On a flat surface, a polygon's angles do not depend on its size. A large regular octagon and a small regular octagon both have inside angles of 135 degrees. On a curved surface, however, the angles do vary with size. On a sphere the angles increase as the polygon grows, whereas on a hyperbolic surface the angles decrease. The above construction requires an octagon that is just the right size to have 45-degree angles, so that when the opposite sides are identified, the eight corners will meet at a single point and the total angle will be 360 degrees. This subtlety explains why the construction would not work with a flat octagon; in Euclidean geometry, eight 135-degree corners cannot meet at a single point. The two-dimensional universe obtained by identifying opposite sides of an octagon must be hyperbolic. The topology dictates the geometry.

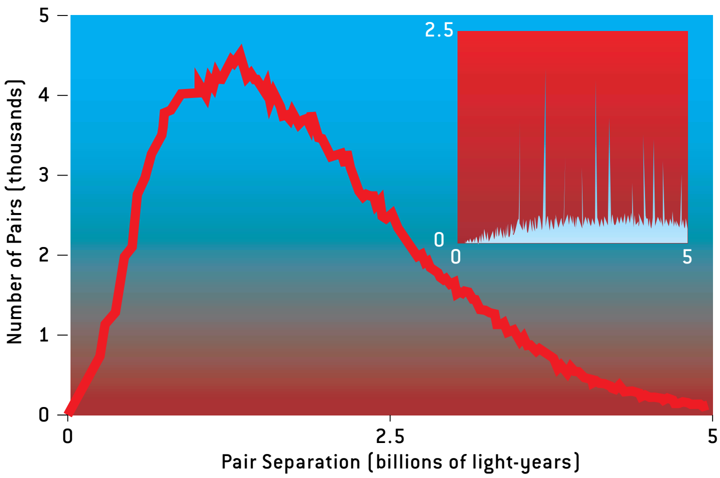
The size of the polygon or polyhedron is measured relative to the only geometrically meaningful length scale for a space: the radius of curvature. A sphere, for example, can have any physical size (in meters, say), but its surface area will always be exactly 4π times the square of its radius—that is, 4π square radians. The same principle applies to the size of a hyperbolic topology, for which a radius of curvature can also be defined. The smallest known hyperbolic space, discovered by one of us (Weeks) in 1985, may be constructed by identifying pairs of faces of an 18-sided polyhedron. It has a volume of approximately 0.94 cubic radian. Other hyperbolic topologies are built from larger polyhedra.

Just as hyperbolic geometry allows for many topologies, so does spherical geometry. In three dimensions, the sphere is generalized into the hypersphere. (To visualize the hypersphere, think of it as being composed of two solid balls in Euclidean space, glued together along their surface: each point of the surface of one ball is the same as the corresponding point on the other ball.) The hypersphere's volume is exactly $2\pi^2$ times the cube of its curvature radius. As early as 1917, Dutch astronomer



FINITE HYPERBOLIC SPACE is formed by an octagon whose opposite sides are connected, so that anything crossing one edge reenters from the opposite edge (*top left*). Topologically, the octagonal space is equivalent to a two-holed pretzel (*top right*). Observers on the surface would see an infinite octagonal grid of galaxies. Such a grid can be drawn only on a hyperbolic manifold, a strange floppy surface where every point has the geometry of a saddle (*bottom*).

SOURCE: JEFFREY R. WEEKS



DISTANCES BETWEEN GALAXY CLUSTERS do not show the pattern expected for a finite, interconnected universe—namely, sharp peaks at distances related to the true size of the cosmos (*inset*). But we (the authors) only studied clusters within roughly two billion light-years of the earth. The universe could still be interconnected on larger scales.

SOURCE: JEAN-PIERRE LUMINET

Willem de Sitter distinguished the projective three-sphere (P^3) from the ordinary three-sphere, S^3. The projective sphere is formed from the sphere by identifying all pairs of antipodal points (those directly opposite each other on the sphere). P^3 therefore has half the volume of S^3. Aside from P^3, there are infinitely many topologies with spherical geometries. Unlike in hyperbolic geometry—in which the more "complicated" the topology, the larger the volume of the fundamental polyhedron—in spherical geometry topological complexity leads to ever smaller fundamental polyhedra. For example, the Poincaré space is represented by a dodecahedron whose opposite faces are pairwise identified; it has a volume $1/120$ that of the hypersphere.

Cosmic space may well have such a shape, in which case an extraordinary "spherical lens" would be generated, with images of cosmic sources repeating according to the Poincaré space's 120-fold "crystal structure." From a mathematical point of view, the volume of the universe, in cubic radians, can be arbitrarily small even if the curvature of the universe is very large. This means that no matter how close to Euclidean space is observed to be, it will always be worth looking for spherical cosmic topology.

Diverse astronomical observations imply that the density of matter in the cosmos is only a third of that needed for space to be Euclidean. Until recently, it was not known whether a cosmological constant made up the difference [see "Cosmological Antigravity," on page 30] or the universe had a hyperbolic geometry with a radius of curvature of 18 billion light-years. Yet recent measurements of the cosmic microwave background strongly suggest that the geometry is at least quite close to Euclidean. Those results, as well as careful measurements of distant supernovae, confirm that something very much like a cosmological constant is prevalent. Nevertheless, many compact topologies—Euclidean, hyperbolic and especially spherical—remain ripe for detection.

The decades from 1930 to 1990 were the dark ages of thinking on cosmic topology. But the 1990s saw the rebirth of the subject. Roughly as many papers have been published on cosmic topology in the past several years as in the preceding 80. Most exciting of all, cosmologists are finally poised to determine the topology observationally.

The simplest test of topology is to look at the arrangement of galaxies. If they lie in a rectangular lattice, with images of the same galaxy repeating at equivalent lattice points, the universe is a 3-torus. Other patterns reveal more complicated topologies. Unfortunately, looking for such patterns can be difficult, because the images of a galaxy would depict different points in its history. Astronomers would need to recognize the same galaxy despite changes in appearance or shifts in position relative to neighboring galaxies. Over the past quarter of a century researchers such as Dmitri Sokoloff of Moscow State University, Viktor Shvartsman of the Soviet Academy of Sciences in Moscow, J. Richard Gott III of Princeton University and Helio V. Fagundes of the Institute for Theoretical Physics in São Paulo have looked for and found no repeating images within one billion light-years of the earth.

Others—such as Boudewijn F. Roukema, now at the Center of Astronomy of the Nicolaus Copernicus University in Torun, Poland—have sought patterns among quasars. Because these objects, thought to be powered by black holes, are bright, any patterns among them can be seen from large distances. The observers identified all groupings of four or more quasars. By examining the spatial relations within each group, they checked whether any pair of groups could in fact be the same group seen from two different directions. Roukema identified two possibilities, but they may not be statistically significant.

Roland Lehoucq and Marc Lachièze-Rey of the department of astrophysics at CEA Saclay in France, together with Jean-Philippe Uzan of the Theoretical Physics Laboratory in Orsay and one of us (Luminet), have circumvented the problems of galaxy recognition in the following way. We have developed various methods of cosmic crystallography that can make out a pattern statistically without needing to recognize specific galaxies as images of one another. If galaxy images repeat periodically, a histogram of all galaxy-to-galaxy distances should show peaks at certain distances, which reflect the true size of the universe. The method has been shown to work well theoretically in a Euclidean or spherical universe. So far we have seen no pat-

THE AUTHORS

JEAN-PIERRE LUMINET, GLENN D. STARKMAN and *JEFFREY R. WEEKS* say they relish participating in the boom years of cosmic topology, as researchers come together across disciplinary boundaries. Luminet, who studies black holes and cosmology at Paris Observatory, has written several books of science as well as poetry and has collaborated with composer Gérard Grisey on the musical performance *Le Noir de l'étoile*. Starkman was institutionalized for six years—at the Institute for Advanced Study in Princeton, N.J., and then at the Canadian Institute for Theoretical Astrophysics in Toronto. He has been released into the custody of Case Western Reserve University. Weeks, the mathematician of the trio, was named a MacArthur Fellow in 1999 and is currently an independent scholar.

tern [*see illustration on preceding page*], but this may be because of the paucity of real data on galaxies farther away than two billion light-years. The Sloan Digital Sky Survey—an ongoing American-Japanese collaboration to prepare a three-dimensional map of much of the universe—and other high-redshift galaxy surveys in progress will produce a larger data set for these studies.

Finally, several other research groups plan to ascertain the topology of the universe using the cosmic microwave background, the faint glow remaining from the big bang. The radiation is remarkably homogeneous: its temperature and intensity are the same in all parts of the sky to nearly one part in 100,000. But there are slight undulations discovered in 1991 by the Cosmic Background Explorer (COBE) satellite. Roughly speaking, the microwave background depicts density variations in the early universe, which ultimately seeded the growth of stars and galaxies.

Circular Reasoning

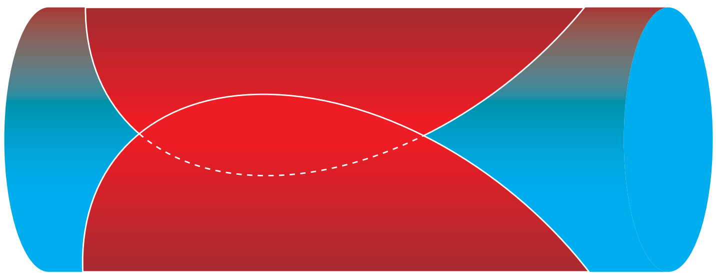
THESE FLUCTUATIONS are the key to resolving a variety of cosmological issues, and topology is one of them. Microwave photons arriving at any given moment began their journeys at approximately the same time and distance from the earth. So their starting points form a sphere, called the last scattering surface, with the earth at the center. Just as a sufficiently large paper disk overlaps itself when wrapped around a broom handle, the last scattering surface will intersect itself if it is big enough to wrap all the way around the universe. The intersection of a sphere with itself is simply a circle of points in space.

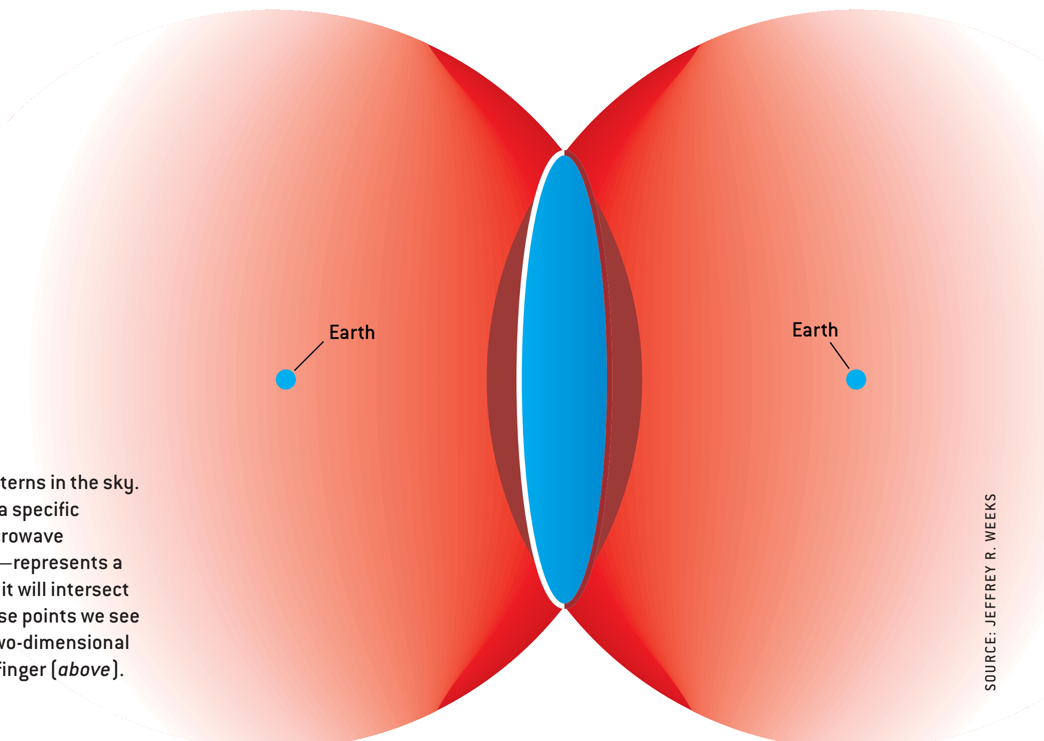
Looking at that circle from the earth, astronomers would see two circles in the sky that share the same pattern of temperature variations. Those two circles are really the same circle in space seen from two perspectives [*see illustration at right*]. They are analogous to the multiple images of a candle in a mirrored room, each of which shows the candle from a different angle.

Two of us (Starkman and Weeks), working with David N. Spergel of Princeton and Neil J. Cornish of Montana State University, hope to detect such circle pairs. The beauty of this method is that it is unaffected by the uncertainties of contemporary cosmology—it relies on the observation that space has constant curvature, but it makes no assumptions about the density of matter, the geometry of space or the presence of a cosmological constant. The main problem is to identify the circles despite the forces that tend to distort their images. For example, as galaxies coalesce, they exert a varying gravitational pull on the radiation as it travels toward the earth, shifting its energy.

Unfortunately, COBE was incapable of resolving structures on an angular scale of less than 10 degrees. Moreover, it did not identify individual hot or cold spots; all one could say for sure is that statistically some of the fluctuations were real features rather than instrumental artifacts. Higher-resolution and lower-noise instruments have since been developed. Some are already making observations from ground-based or balloon-borne observatories, but they do not cover the whole sky. The crucial observations will be made by NASA's Microwave Anisotropy Probe (MAP), now gathering data, and the European Space Agency's Planck satellite, scheduled for launch in 2007.

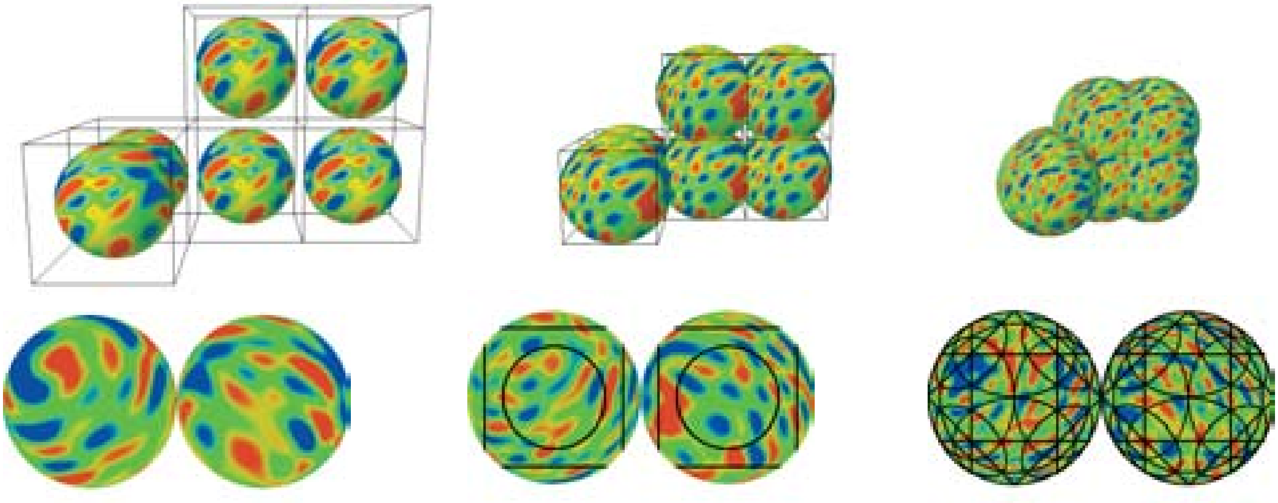
The relative positions of the matching circles, if any, will reveal the specific topology of the universe. If the last scattering surface is barely big enough to wrap around the universe, it will





WRAPPED AROUND THE COSMOS, light creates patterns in the sky. All the light received from a specific time or from a specific distance from the earth—such as the cosmic microwave background radiation left over from the big bang—represents a sphere. If this sphere is larger than the universe, it will intersect itself, defining a circle. This circle consists of those points we see twice: from the left and from the right (*right*). A two-dimensional analogy is a circular bandage wrapped around a finger (*above*).

SOURCE: JEFFREY R. WEEKS



THREE POSSIBLE UNIVERSES, large, medium and small (*top row*), would produce distinctive patterns in the cosmic microwave background radiation, as simulated here (*bottom row*). Each of these universes has the topology of a 3-torus and is shown repeated six times to evoke the regular grid that an observer would see. In the large universe, the sphere of background radiation does not overlap itself, so no patterns emerge. In the medium universe, the sphere intersects itself once in each direction. One may verify that tracing clockwise around the central circle in the left hemisphere reveals the same sequence of colors as tracing counter-clockwise in the right. Finally, in the small universe, the sphere intersects itself many times, resulting in a more complex pattern.

JEFFREY R. WEEKS

intersect only its nearest ghost images. If it is larger, it will reach farther and intersect the next nearest images. If the last scattering surface is large enough, we expect hundreds or even thousands of circle pairs [*see illustration above*]. The data will be highly redundant. The largest circles will completely determine the topology of space as well as the position and orientation of all smaller circle pairs. Thus, the internal consistency of the patterns will verify not just the correctness of the topological findings but also the correctness of the microwave background data.

Other teams have different plans for the data. John D. Barrow and Janna J. Levin of the University of Cambridge, Emory F. Bunn of St. Cloud State University, Evan Scannapieco of the Astrophysical Observatory of Arcetri, Italy, and Joseph I. Silk of the University of Oxford intend to examine the pattern of hot and cold spots directly. The group has already constructed sample maps simulating the microwave background for particular topologies. They have multiplied the temperature in each direction by the temperature in every other direction, generating a huge four-dimensional map of what is usually called the two-point correlation function. The maps provide a quantitative way of comparing topologies. J. Richard Bond of the Canadian Institute for Theoretical Astrophysics in Toronto, Dmitry Pogosyan of the University of Alberta and Tarun Souradeep of the Inter-University Center for Astronomy and Astrophysics in Pune, India, were among the first to apply related new techniques to the existing COBE data, which could accurately identify the smallest hyperbolic spaces.

Beyond the immediate intellectual satisfaction, discovering the topology of space would have profound implications for physics. Although relativity says nothing about the universe's topology, newer and more comprehensive theories that are under development should predict the topology or at least assign probabilities to the various possibilities. These theories are needed to explain gravity in the earliest moments of the big bang, when quantum-mechanical effects were important [see "Quantum Gravity," by Bryce S. DeWitt; SCIENTIFIC AMERICAN, December 1983]. The theories of everything, such as M-theory, are in their infancy and do not yet have testable consequences. But eventually the candidate theories will make predictions about the topology of the universe on large scales.

The tentative steps toward the unification of physics have already spawned the subfield of quantum cosmology. There are three basic hypotheses for the birth of the universe, which are advocated, respectively, by Andrei Linde of Stanford University, Alexander Vilenkin of Tufts University and Stephen W. Hawking of the University of Cambridge. One salient point of difference is whether the expected volume of a newborn universe is very small (Linde's and Vilenkin's proposals) or very large (Hawking's). Topological data may be able to distinguish among these models.

Since ancient times, cultures around the world have asked how the universe began and whether it is finite or infinite. Through a combination of mathematical insight and careful observation, science in the 20th century partially answered the first question. It might begin the 21st century with an answer to the second as well. SA

MORE TO EXPLORE

La Biblioteca de Babel (The Library of Babel). Jorge Luis Borges in *Ficciones.* Emecé Editores, 1956. Text available at **jubal.westnet.com/hyperdiscordia/library_of_babel.html**

Cosmic Topology. Marc Lachièze-Rey and Jean-Pierre Luminet in *Physics Reports,* Vol. 254, No. 3, pages 135–214; March 1995. Preprint available at **xxx.lanl.gov/abs/gr-qc/9605010**

Circles in the Sky: Finding Topology with the Microwave Background Radiation. Neil J. Cornish, David N. Spergel and Glenn D. Starkman in *Classical and Quantum Gravity,* Vol. 15, No. 9, pages 2657–2670; September 1998. Preprint available at **xxx.lanl.gov/abs/astro-ph/9801212**

Reconstructing the Global Topology of the Universe from the Cosmic Microwave Background. Jeffrey R. Weeks in *Classical and Quantum Gravity,* Vol. 15, No. 9, pages 2599–2604; September 1998. Preprint available at **xxx.lanl.gov/abs/astro-ph/9802012**

Free software for exploring topology is available at **www.northnet.org/weeks**

L'Univers chiffonné. Jean-Pierre Luminet. Fayard, Paris, 2001.

The Shape of Space. Jeffrey R. Weeks. Marcel Dekker, Paris, 2002.

STRUCTURE

The visible universe could lie on a membrane floating within a higher-dimensional space

the Universe's unseen DIMENSIONS

BY NIMA ARKANI-HAMED, SAVAS DIMOPOULOS AND GEORGI DVALI

ILLUSTRATIONS BY BRYAN CHRISTIE DESIGN

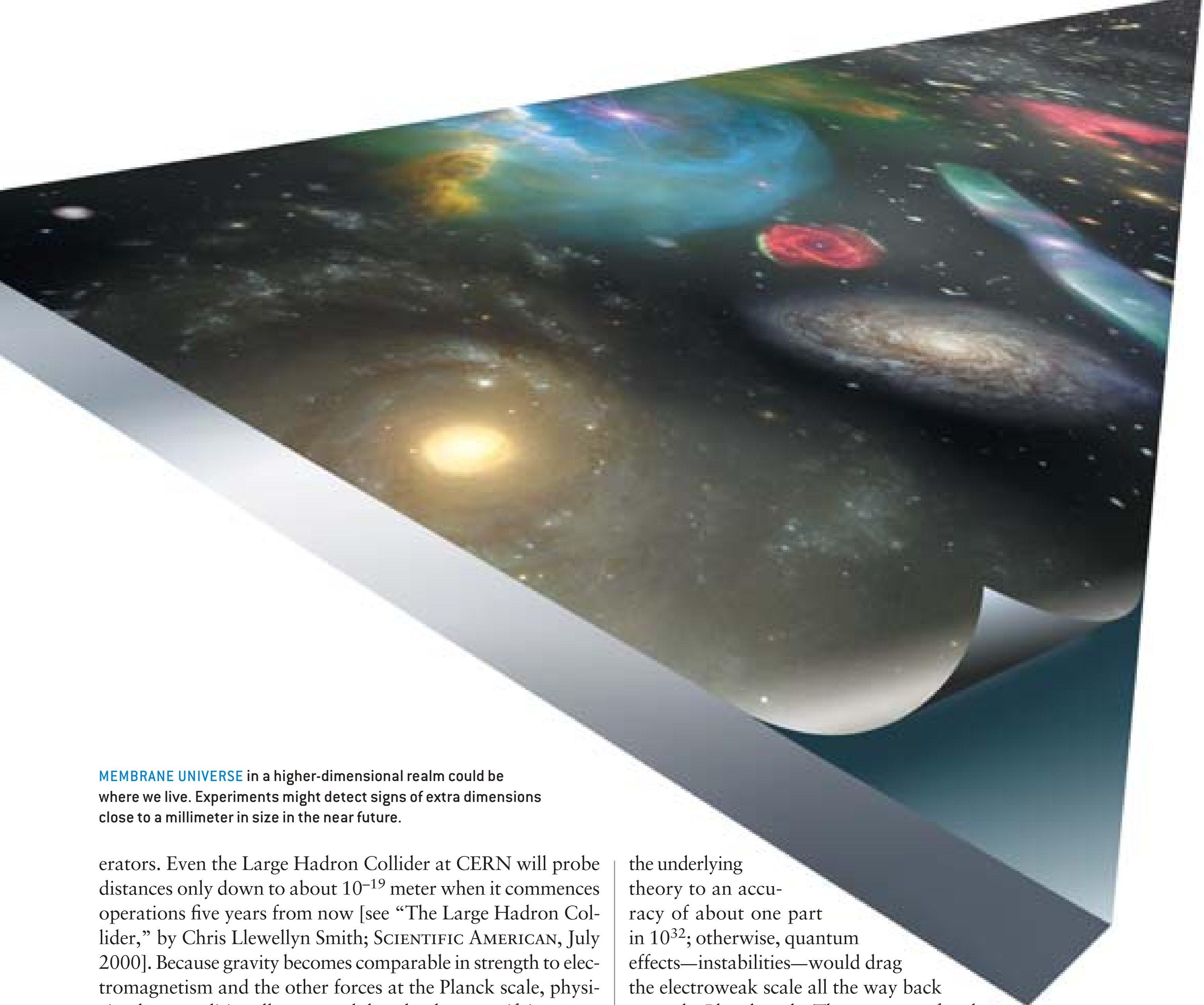
THE CLASSIC 1884 story *Flatland: A Romance of Many Dimensions,* by Edwin A. Abbott, describes the adventures of "A. Square," a character who lives in a two-dimensional world populated by animated geometric figures—triangles, squares, pentagons and so on. Toward the end of the story, on the first day of 2000, a spherical creature from three-dimensional "Spaceland" passes through Flatland and carries A. Square up off his planar domain to show him the true three-dimensional nature of the larger world. As he comes to grasp what the sphere is showing him, A. Square speculates that Spaceland may itself exist as a small subspace of a still larger four-dimensional universe.

Amazingly, in the past four years physicists have begun seriously examining a very similar idea: that everything we can see in our universe is confined to a three-dimensional "membrane" that lies within a higher-dimensional realm. But unlike A. Square, who had to rely on divine intervention from Spaceland for his insights, physicists may soon be able to detect and verify the existence of reality's extra dimensions, which could extend over distances as large as a millimeter. Experiments are already looking for the extra dimensions' effect on the force of gravity. If the theory is correct, upcoming high-energy particle experiments in Europe could see unusual processes involving quantum gravity, such as the creation of transitory micro black holes. More than just an idle romance of many dimensions, the theory is based on some of the most recent developments in string theory and would solve some long-standing puzzles of particle physics and cosmology.

The exotic concepts of string theory and multidimensions actually arise from attempts to understand the most familiar of forces: gravity. More than three centuries after Isaac Newton proposed his law of gravitation, physics still does not explain why gravity is so much weaker than all the other forces. The feebleness of gravity is dramatic. A small magnet readily overcomes the gravitational pull of the entire mass of the earth when it lifts a nail off the ground. The gravitational attraction between two electrons is 10^{43} times weaker than the repulsive electric force between them. Gravity seems important to us—keeping our feet on the ground and the earth orbiting the sun—only because these large aggregates of matter are electrically neutral, making the electrical forces vanishingly small and leaving gravity, weak as it is, as the only noticeable force left over.

The Inexplicable Weakness of Gravity

ELECTRONS WOULD HAVE to be 10^{22} times more massive for the electric and gravitational forces between two of them to be equal. To produce such a heavy particle would take 10^{19} giga-electron volts (GeV) of energy, a quantity known as the Planck energy (after German physicist Max Planck). A related quantity is the Planck length, a tiny 10^{-35} meter. By comparison, the nucleus of a hydrogen atom, a proton, is about 10^{19} times as large and has a mass of about 1 GeV. The Planck scale of energy and length is far out of reach of the most powerful accel-



MEMBRANE UNIVERSE in a higher-dimensional realm could be where we live. Experiments might detect signs of extra dimensions close to a millimeter in size in the near future.

erators. Even the Large Hadron Collider at CERN will probe distances only down to about 10^{-19} meter when it commences operations five years from now [see "The Large Hadron Collider," by Chris Llewellyn Smith; SCIENTIFIC AMERICAN, July 2000]. Because gravity becomes comparable in strength to electromagnetism and the other forces at the Planck scale, physicists have traditionally assumed that the theory unifying gravity with the other interactions would reveal itself only at these energies. The nature of the ultimate unified theory would then be hopelessly out of reach of direct experimental investigation in the foreseeable future [see "A Unified Physics by 2050?" by Steven Weinberg; SCIENTIFIC AMERICAN, December 1999].

Today's most powerful accelerators probe the energy realm between 100 and 1,000 GeV (one teraelectron volt, or TeV). In this range, experimenters have seen the electromagnetic force and the weak interaction (a force between subatomic particles responsible for certain types of radioactive decay) become unified. We would understand gravity's extraordinary weakness if we understood the factor of 10^{16} that separates the electroweak scale from the Planck scale.

Alas, physicists' extremely successful theory of particle physics, called the Standard Model, cannot explain the size of this huge gap, because the theory is carefully adjusted to fit the observed electroweak scale. The good news is that this adjustment (along with about 16 others) serves once and for all to fit myriad observations. The bad news is that we must fine-tune the underlying theory to an accuracy of about one part in 10^{32}; otherwise, quantum effects—instabilities—would drag the electroweak scale all the way back up to the Planck scale. The presence of such delicate balancing in the theory is like walking into a room and finding a pencil standing perfectly on its tip in the middle of a table. Though not impossible, the situation is highly unstable, and we are left wondering how it came about.

For 20 years, theorists have attacked this conundrum, called the hierarchy problem, by altering the nature of particle physics near 10^{-19} meter (or 1 TeV) to stabilize the electroweak scale. The most popular modification of the Standard Model that achieves this goal involves a new symmetry called supersymmetry. Going back to our pencil analogy, supersymmetry acts like an invisible thread holding up the pencil and preventing it from falling over. Although accelerators have not yet turned up any direct evidence for supersymmetry, some suggestive indirect evidence supports the theory. For example, when the measured strengths of the strong, weak and electromagnetic forces are theoretically extrapolated to shorter distances, they meet very accurately at a common value only if supersymmetric rules govern the extrapolation. This result hints at a supersymmetric unification of these three forces at about

10^{-32} meter, approximately 1,000 times larger than the Planck length but still far beyond the range of particle colliders.

Gravity and Large Spatial Dimensions

FOR TWO DECADES, the only viable framework for tackling the hierarchy problem has been to change particle physics near 10^{-19} meter by introducing new processes such as supersymmetry. But in the past four years, theorists have proposed a radically different approach, modifying spacetime, gravity and the Planck scale itself. The key insight is that the extraordinary size of the Planck scale, accepted for a century since Planck first introduced it, is based on an untested assumption about how gravity behaves over short distances.

Newton's inverse square law of gravity—which says the force between two masses falls as the square of the distance between them—works extremely well over macroscopic distances, explaining the earth's orbit around the sun, and so on. But because gravity is so weak, the law has been experimentally tested down to distances of only about a millimeter, and we must extrapolate across 32 orders of magnitude to conclude that gravity becomes strong only at a Planck scale of 10^{-35} meter.

The inverse square law is natural in three-dimensional space [*see upper illustration on opposite page*]. Consider lines of gravitational force emanating uniformly from the earth. Farther from the earth, the lines are spread over a spherical shell of greater area. The surface area increases as the square of the distance, and so the force is diluted at that rate. Suppose there were one more dimension, making space four-dimensional. Then the field lines emanating from a point would get spread over a four-dimensional shell whose surface would increase as the cube of the distance, and gravity would follow an inverse cube law.

The inverse cube law certainly doesn't describe our universe, but now imagine that the extra dimension is curled up into a small circle of radius R and that we're looking at field lines coming from a tiny point mass [*see lower illustration on opposite page*]. When the field lines are much closer to the mass than the distance R, they can spread uniformly in all four dimensions, and so the force of gravity falls as the inverse cube of distance. Once the lines have spread fully around the circle, however, only three dimensions remain for them to continue spreading

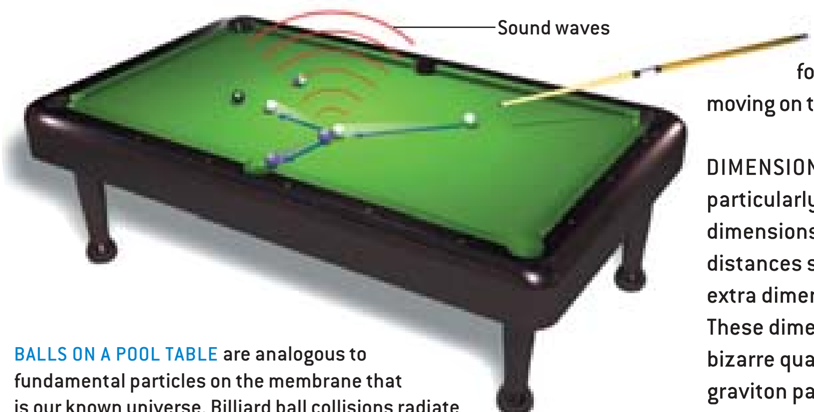
IN A NUTSHELL by Graham P. Collins

DIMENSIONS. Our universe seems to have four dimensions: three of space (up-down, left-right, forward-backward) and one of time. Although we can barely imagine additional dimensions, mathematicians and physicists have long analyzed the properties of theoretical spaces that have any number.

SIZE OF DIMENSIONS. The four known spacetime dimensions of our universe are vast. The dimension of time extends back at least 13 billion years into the past and may extend infinitely into the future. The three spatial dimensions may be infinite; our telescopes have detected objects more than 12 billion light-years away. Dimensions can also be finite. For example, the two dimensions of the surface of the earth extend only about 40,000 kilometers—the length of a great circle.

SMALL EXTRA DIMENSIONS. Some modern physics theories postulate additional real dimensions that are wrapped up in circles so small (perhaps 10^{-35}-meter radius) that we have not detected them. Think of a thread of cotton: to a good approximation, it is one-dimensional. A single number can specify where an ant stands on the thread. But using a microscope, we see dust mites crawling on the thread's two-dimensional surface: along the large length dimension and around the short circumference dimension.

LARGE EXTRA DIMENSIONS. Recently physicists realized that extra dimensions as big as a millimeter could exist and remain invisible to us. Surprisingly, no known experimental data rule out the theory, and it could explain several mysteries of particle physics and cosmology. We and all the contents of our known three-dimensional universe (except for gravity) would be stuck on a membrane, like balls moving on the two-dimensional green baize of a pool table.



BALLS ON A POOL TABLE are analogous to fundamental particles on the membrane that is our known universe. Billiard ball collisions radiate energy into three dimensions as sound waves (*red*), analogous to gravitons. Precise studies of the balls' motions could detect the missing energy and thus the higher dimensions.

DIMENSIONS AND GRAVITY. The behavior of gravity—particularly its strength—is intimately related to how many dimensions it pervades. Studies of gravity acting over distances smaller than a millimeter could thus reveal large extra dimensions to us. Such experiments are under way. These dimensions would also enhance the production of bizarre quantum gravity objects such as micro black holes, graviton particles and superstrings, all of which could be detected sometime this decade at high-energy particle accelerators.

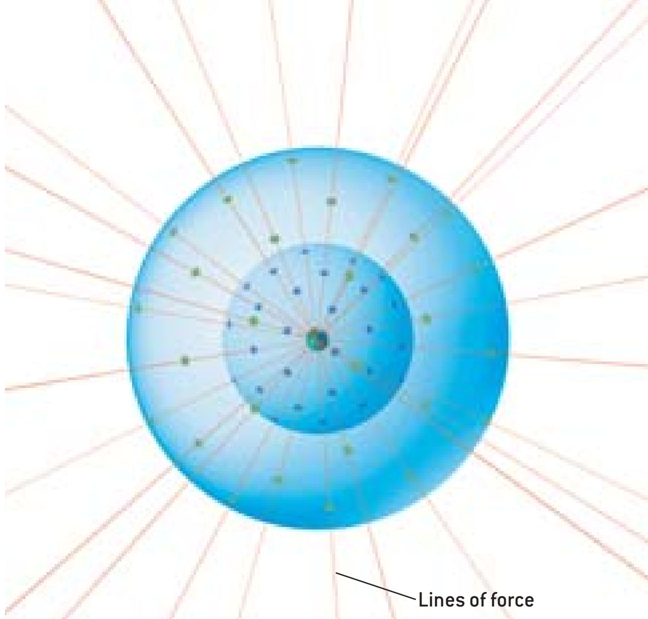
Graham P. Collins is a staff editor and writer.

through, and so for distances much greater than *R* the force varies as the inverse square of the distance.

The same effect occurs if there are many extra dimensions, all curled up into circles of radius *R*. For *n* extra spatial dimensions at distances smaller than *R,* the force of gravity will follow an inverse $2 + n$ power law. Because we have measured gravity only down to about a millimeter, we would be oblivious to changes in gravity caused by extra dimensions for which *R* was smaller than a millimeter. Furthermore, the $2 + n$ power law would cause gravity to reach Planck-scale strength well above 10^{-35} meter. That is, the Planck length (defined by where gravity becomes strong) would not be that small, and the hierarchy problem would be reduced.

One can solve the hierarchy problem completely by postulating enough extra dimensions to move the Planck scale very close to the electroweak scale. The ultimate unification of gravity with the other forces would then take place near 10^{-19} meter rather than 10^{-35} meter as traditionally assumed. How many dimensions are needed depends on how large they are. Conversely, for a given number of extra dimensions we can compute how large they must be to make gravity strong near 10^{-19} meter. If there is only one extra dimension, its radius *R* must be roughly the distance between the earth and the sun. Therefore, this case is already excluded by observation. Two extra dimensions, however, can solve the hierarchy problem if they are about a millimeter in size—precisely where our direct knowledge of gravity ends. The dimensions are smaller still if we add more of them, and for seven extra dimensions we need them to be around 10^{-14} meter big, about the size of a uranium nucleus. This is tiny by everyday standards but huge by the yardstick of particle physics.

Postulating extra dimensions may seem bizarre and ad hoc, but to physicists it is an old, familiar idea that dates back to the 1920s, when Polish mathematician Theodor Kaluza and Swedish physicist Oskar Klein developed a remarkable unified theory of gravity and electromagnetism that required one extra dimension. The idea has been revived in modern string theories, which require a total of 10 spatial dimensions for internal mathematical consistency. In the past, physicists have assumed that the extra dimensions are curled up into tiny circles with a size



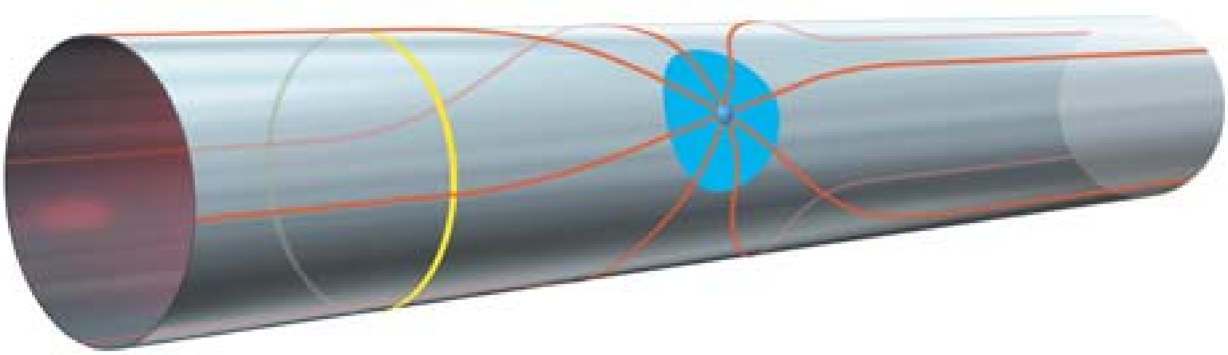
GRAVITATIONAL LINES OF FORCE spread out from the earth in three dimensions. As distance from the earth increases, the force becomes diluted by being spread across a larger surface area (*spheres*). The surface area of each sphere increases as the square of its radius, so gravity falls as the inverse square of distance in three dimensions.

near the traditional Planck length of 10^{-35} meter, making them undetectable but also leaving the conundrum of the hierarchy problem. In contrast, in the new theory that we are discussing, the extra dimensions are wrapped into big circles of at least 10^{-14} meter radius and perhaps as enormous as a millimeter.

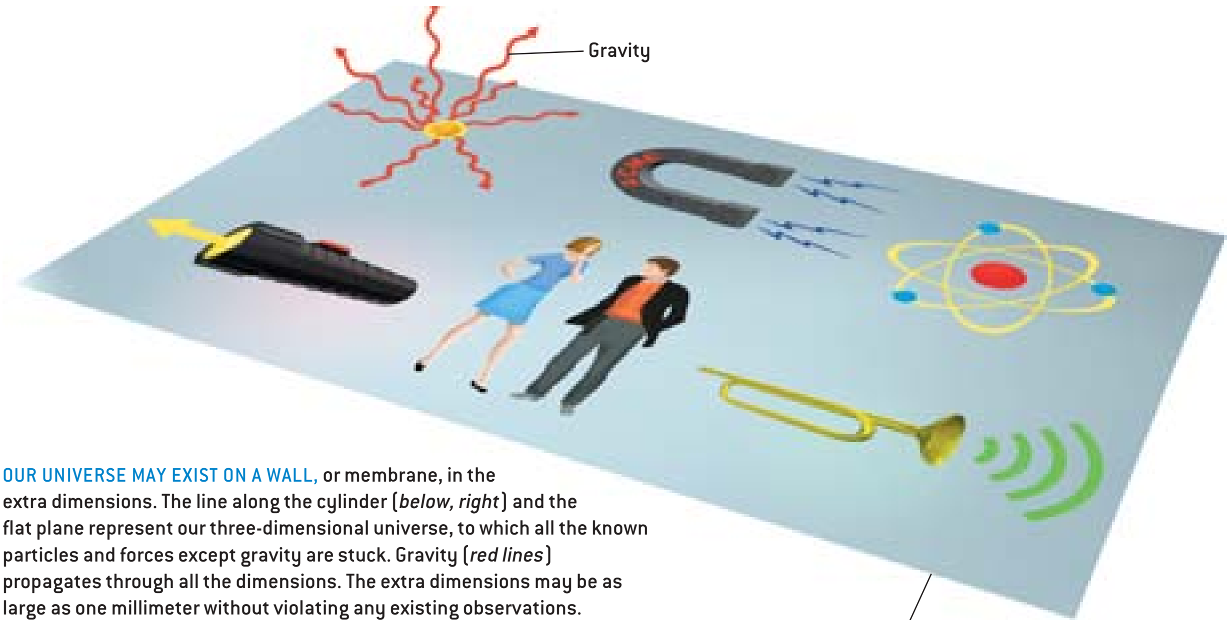
Our Universe on a Wall

IF THESE DIMENSIONS are that large, why haven't we seen them yet? Extra dimensions a millimeter big would be discernible to the naked eye and obvious through a microscope. And although we have not measured gravity much below about a millimeter, we have a wealth of experimental knowledge concerning all the other forces at far shorter distances, approaching 10^{-19} meter, all of it consistent only with three-dimensional space. How could there possibly be large extra dimensions?

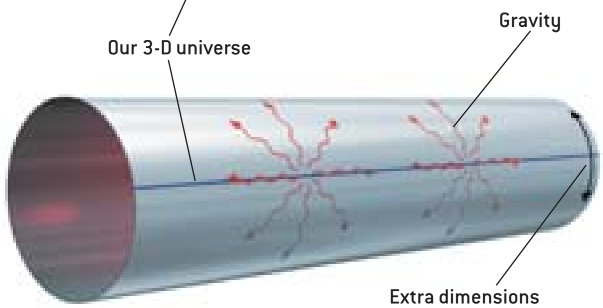
The answer is at once simple and peculiar: all the matter and forces we know of—with the sole exception of gravity—are stuck to a "wall" in the space of the extra dimensions [*see illustration on next page*]. Electrons, protons, photons and all the other particles in the Standard Model cannot move in the extra



SMALL EXTRA DIMENSION wrapped in a circle (*circumference of tube*) modifies how gravity (*red lines*) spreads in space. At distances smaller than the circle radius (*blue patch*), the lines of force spread apart rapidly through all the dimensions. At much larger distances (*yellow circle*), the lines have filled the extra dimension, which has no further effect on the lines of force.



OUR UNIVERSE MAY EXIST ON A WALL, or membrane, in the extra dimensions. The line along the cylinder (*below, right*) and the flat plane represent our three-dimensional universe, to which all the known particles and forces except gravity are stuck. Gravity (*red lines*) propagates through all the dimensions. The extra dimensions may be as large as one millimeter without violating any existing observations.



dimensions; electric and magnetic field lines cannot spread into the higher-dimensional space. The wall has only three dimensions, and as far as these particles are concerned, the universe might as well be three-dimensional. Only gravitational field lines can extend into the higher-dimensional space, and only the particle that transmits gravity, the graviton, can travel freely into the extra dimensions. The presence of the extra dimensions can be felt only through gravity.

To make an analogy, imagine that all the particles in the Standard Model, like electrons and protons, are billiard balls moving on the surface of a vast pool table. As far as they are concerned, the universe is two-dimensional. Nevertheless, pool-table inhabitants made out of billiard balls could still detect the higher-dimensional world: when two balls hit each other sufficiently hard, they produce sound waves, which travel in all three dimensions, carrying some energy away from the table surface [*see illustration on page 68*]. The sound waves are analogous to gravitons, which can travel in the full higher-dimensional space. In high-energy particle collisions, we expect to observe missing energy, the result of gravitons escaping into the extra dimensions.

Although it may seem strange that some particles should be confined to a wall, similar phenomena are quite familiar. For instance, electrons in a copper wire can move only along the one-dimensional space of the wire and do not travel into the surrounding three-dimensional space. Likewise, water waves travel primarily on the surface of the ocean, not throughout its depth. The specific scenario we are describing, in which all particles except gravity are stuck to a wall, can arise naturally in string theory. In fact, one of the major insights triggering recent breakthroughs in string theory has been the recognition that the theory contains such walls, known as D-branes ("brane" comes from the word "membrane," and "D" stands for "Dirichlet," which indicates a mathematical property of the branes). D-branes have precisely the required features: particles such as electrons and photons are represented by tiny lengths of string that each have two end points that must be stuck to a D-brane. Gravitons, on the other hand, are tiny closed loops of string that can wander into all the dimensions because they have no end points anchoring them to a D-brane.

Is It Alive?

ONE OF THE FIRST THINGS good theorists do when they have a new theory is to try to kill it by finding an inconsistency with experimental results. The theory of large extra dimensions changes gravity at macroscopic distances and alters other physics at high energies, so surely it is easy to kill. Remarkably, however, it does not contradict any known experiment. A few examples show how surprising this conclusion is.

One might initially worry that changing gravity would affect objects held together by gravity, such as stars and galaxies. But they are not affected. Gravity changes only at distances shorter than a millimeter, whereas in a star, for example, gravity acts across thousands of kilometers to hold distant parts of the star together.

A much more serious concern relates to gravitons, the hypothetical particles that transmit gravity in a quantum theory. In the theory with extra dimensions, gravitons interact much more strongly with matter, so many more of them should be produced in high-energy particle collisions. In addition, they propagate in all the dimensions, thus taking energy away from the wall, or membrane, that is the universe where we live.

When a star collapses and explodes as a supernova, the high temperatures can readily boil off gravitons into extra dimensions [*see upper illustration on next page*]. From observations of the famous supernova 1987A, however, we know that the explosion emits most of its energy as neutrinos, leaving little room for any energy leakage by gravitons. Our understanding of supernovae therefore limits how strongly gravitons can couple to matter. This constraint could easily have killed the idea of large extra dimensions, but detailed calculations show that the theory survives. The most severe limit is for only two extra dimensions, in which case gravitons cool supernovae too much.

Theorists have examined many other possible constraints based on unacceptable changes. The theory passes all these experimental checks, which turn out to be less stringent than the supernova constraint. Perhaps surprisingly, the constraints become less severe as more dimensions are added to the theory. We saw this right from the start: the case of one extra dimension was excluded immediately because gravity would be altered at solar system distances. This indicates why more dimensions are safer: the dramatic strengthening of gravity begins at shorter distances and therefore has a smaller impact on the larger-distance processes.

Answers by 2010

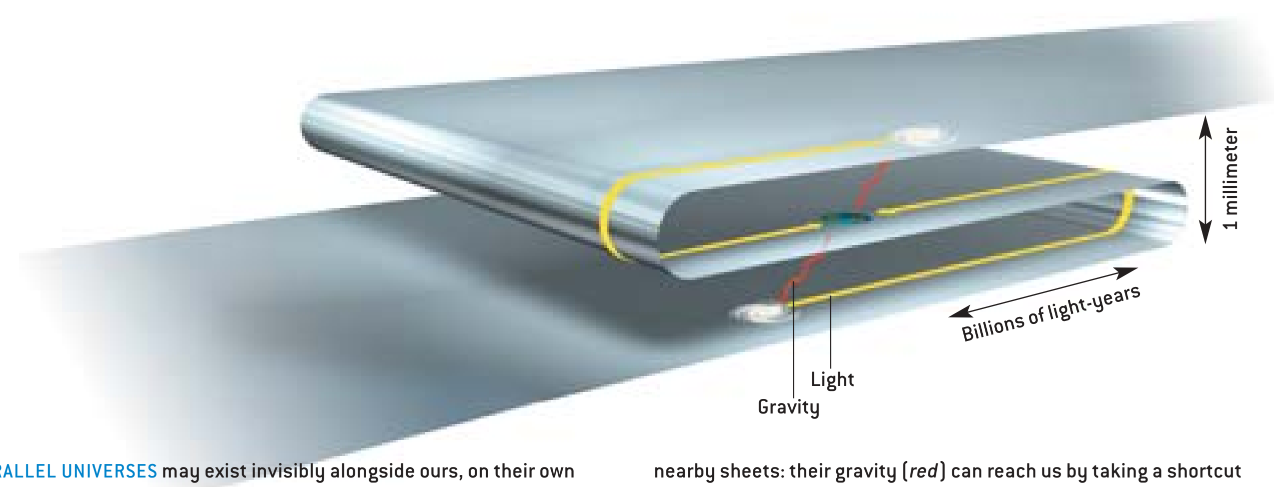
THE THEORY SOLVES the hierarchy problem by making gravity a strong force near TeV energies, precisely the energy scale to be probed using upcoming particle accelerators. Experiments at the Large Hadron Collider (LHC), due to begin around 2007, should therefore uncover the nature of quantum gravity! For instance, if string theory is the correct description of quantum gravity, particles are like tiny loops of string that can vibrate like a violin string. The known fundamental particles correspond to a string that is not vibrating, much like an unbowed violin string. Each different "musical note" that a string can carry by vibrating would appear as a different, exotic new particle. In conventional string theories, the strings have been thought of as only 10^{-35} meter long, and the new particles would have masses on the order of the traditional Planck energy—the "music" of such strings would be too high-pitched for us to "hear" at particle colliders. But with large extra dimensions, the strings are much longer, near 10^{-19} meter, and the new particles would appear at TeV energies—low enough to hear at the LHC.

Similarly, the energies needed to create micro black holes in particle collisions would fall within experimental range [*see lower illustration on next page*].

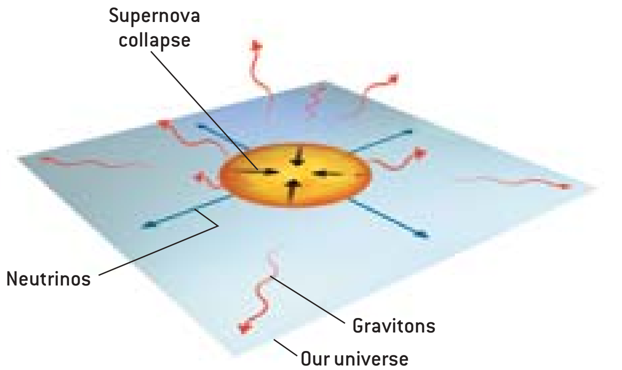
Even at energies too low to produce vibrating strings or black holes, particle collisions would produce large numbers of gravitons, a process that is negligible in conventional theories. The experiments could not directly detect the emitted gravitons, but the energy they carry off would show up as energy missing from the collision debris. The theory predicts specific properties of the missing energy—how it should vary with collision energy and so on—so evidence of graviton production can be distinguished from other processes that can carry off energy in unseen particles. Current data from the highest-energy accelerators already mildly constrain the large-dimensions scenario. Experiments at the LHC should either see evidence of gravitons or begin to exclude the theory by their absence.

A completely different type of experiment could also substantiate the theory, perhaps much sooner than the particle colliders. Recall that for two extra dimensions to solve the hierarchy problem, they must be as large as a millimeter. Measurements of gravity would then detect a change from Newton's inverse square law to an inverse fourth power law at distances near a millimeter. Extensions of the basic theoretical framework lead to a whole host of other possible deviations from Newtonian gravity, the most interesting of which is *repulsive* forces more than a million times stronger than gravity occurring between masses separated by less than a millimeter. Tabletop experiments using exquisitely built detectors are now under way, testing Newton's law from the centimeter range down to tens of microns [*see illustration on page 73*].

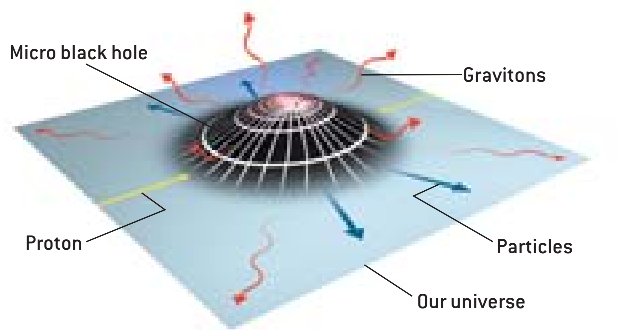
To probe the gravitational force at submillimeter distances, one must use objects not much larger than a millimeter, which therefore have very small masses. One must carefully screen out numerous effects, such as residual electrostatic forces, that could



PARALLEL UNIVERSES may exist invisibly alongside ours, on their own membranes less than a millimeter away from ours. Such parallel universes could also be different sheets of our own universe folded back on itself. So-called dark matter could be explained by ordinary stars and galaxies on nearby sheets: their gravity (*red*) can reach us by taking a shortcut through the extra dimensions, but we cannot see them because light (*yellow*) must travel billions of light-years to the folds and back before it reaches the earth.



SUPERNOVA occurs when the collapse of a massive star produces an explosive shock wave. Most of the energy is emitted as neutrinos (*blue*). If extra dimensions exist, radiated gravitons (*red*) carry away more energy than they would in three dimensions. Theorists constrain the properties of the extra dimensions by requiring that energy leakage by gravitons not cause supernovae to fizzle.



MICRO BLACK HOLES could be created in particle accelerators such as the Large Hadron Collider by smashing together protons (*yellow*) at high energies. The holes would evaporate rapidly by emitting Hawking radiation of Standard Model particles (*blue*) and gravitons (*red*).

mask or fake the tiny gravitational attraction. Such experiments are difficult and subtle, but it is exciting that they might uncover dramatic new physics. Even apart from the search for extra dimensions, it is important to extend our direct knowledge of gravity to these short distances. Researchers at the University of Washington have performed a measurement of gravity down to one fifth of a millimeter and have found no deviations from Newtonian gravity. Therefore, any large new dimensions must be less than a fifth of a millimeter in size. Several groups are now looking to improve on this measurement.

The idea of extra dimensions in effect continues the Copernican tradition in understanding our place in the world: The earth is not the center of the solar system, the sun is not the center of our galaxy, our galaxy is just one of billions in a universe that has no center, and now our entire three-dimensional universe would be just a thin membrane in the full space of dimensions. If we consider slices across the extra dimensions, our universe would occupy a single infinitesimal point in each slice, surrounded by a void.

Perhaps this is not the full story. Just as the Milky Way is not the only galaxy in the universe, might our universe not be alone in the extra dimensions? The membranes of other three-dimensional universes could lie parallel to our own, only a millimeter removed from us in the extra dimensions [*see illustration on preceding page*]. Similarly, although all the particles of the Standard Model must stick to our own membrane universe, other particles beyond the Standard Model might propagate through the extra dimensions. Far from being empty, the extra dimensions could have a multitude of interesting structures.

The effects of new particles and universes in the extra dimensions may provide answers to many outstanding mysteries of particle physics and cosmology. For example, they may account for the masses of the ghostly elementary particles called neutrinos. Impressive evidence from the Super Kamiokande experiment in Japan indicates that neutrinos, long assumed to be massless, have a minuscule but nonzero mass. The neutrino can gain its mass by interacting with a partner field living in the extra dimensions. As with gravity, the interaction is greatly diluted by the partner's being spread throughout the extra dimensions, and so the neutrino acquires only a tiny mass.

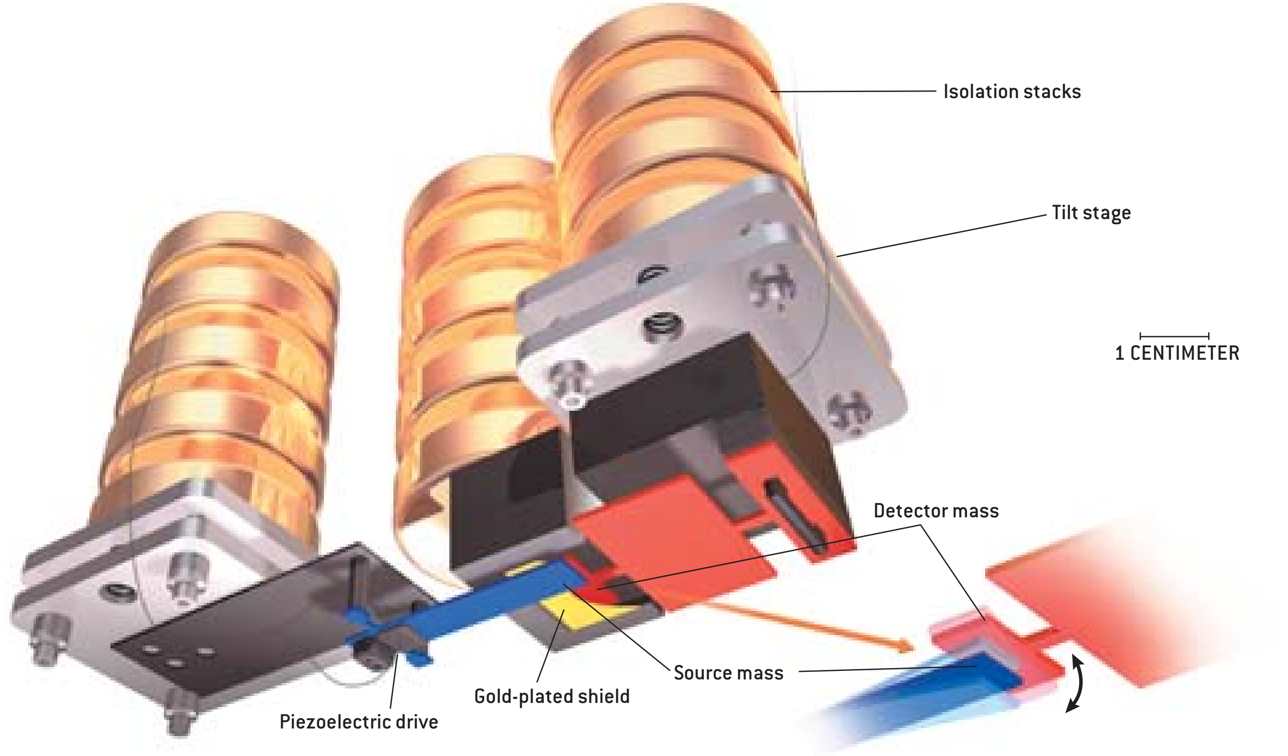
Parallel Universes

ANOTHER EXAMPLE is the mystery in cosmology of what constitutes dark matter, the invisible gravitating substance that seems to make up more than 90 percent of the mass of the universe. Dark matter may reside in parallel universes. Such matter would affect our universe through gravity and is necessarily "dark" because our species of photon is stuck to our membrane, so photons cannot travel across the void from the parallel matter to our eyes.

Such parallel universes might be utterly unlike our own, having different particles and forces and perhaps even being confined to membranes with fewer or more dimensions. In one intriguing scenario, however, they have identical properties to our own world. Imagine that the wall where we live is folded a number of times in the extra dimensions [*see illustration on preceding page*]. Objects on the other side of a fold will appear to be very distant even if they are less than a millimeter from us in the extra dimensions: the light they emit must travel to the crease and back to reach us. If the crease is tens of billions of light-years away, no light from the other side could have reached us since the universe began.

THE AUTHORS

NIMA ARKANI-HAMED, SAVAS DIMOPOULOS and *GEORGI DVALI* conceived the extra-dimension theory at Stanford University in February 1998. Arkani-Hamed was born in Houston and in 1997 received a Ph.D. in physics at the University of California, Berkeley, where he has been assistant professor since 1999. When he's not exploring theoretical possibilities beyond the Standard Model of particle physics, he enjoys hiking in the High Sierra and the California desert. Dimopoulos grew up in Athens, received a Ph.D. from the University of Chicago and has been professor of physics at Stanford since 1979. His research has mostly been driven by the quest for what lies beyond the Standard Model. In 1981, together with Howard Georgi of Harvard University, he proposed the supersymmetric Standard Model. "Gia" Dvali was raised in what is now the Republic of Georgia and in 1992 received his Ph.D. in high-energy physics and cosmology from Tbilisi State University. In 1998 he became associate professor of physics at New York University. He enjoys overcoming gravity by high mountaineering and rock and ice climbing.



TORSION OSCILLATOR at the University of Colorado looks for changes in gravity from 0.05 to 1.0 millimeter. Piezoelectrics vibrate the tungsten source mass (*blue*) like a diving board. Any forces acting between the source mass and the tungsten detector (*red*) produce twisting oscillations of the detector (*inset; oscillations are exaggerated*), which are sensed by electronics. A gold-plated shield (*yellow*) suppresses electrostatic forces, and suspension from brass isolation stacks stops vibrations from traveling from the source to the detector. Electrostatic shields enclosing the apparatus are not shown. For maximum sensitivity, liquid helium cools the apparatus to four kelvins.

Dark matter could be composed of ordinary matter, perhaps even ordinary stars and galaxies, shining brightly on their own folds. Such stars would produce interesting observable effects, such as gravitational waves from supernovae. Gravitational-wave detectors scheduled for completion soon could find evidence for folds by observing large sources of gravitational radiation that cannot be accounted for by matter visible in our own universe.

Our theory is not the first proposal involving extra dimensions larger than 10^{-35} meter. In 1990 Ignatios Antoniadis of the École Polytechnique in France suggested that some of string theory's dimensions might be as large as 10^{-19} meter. In 1996 Petr Hořava of the California Institute of Technology and Edward Witten of the Institute for Advanced Study in Princeton, N.J., pointed out that a single extra dimension of 10^{-30} meter would neatly unify forces. Following this idea, Joseph Lykken of Fermi National Accelerator Laboratory in Batavia, Ill., attempted to lower the unification scale to near 10^{-19} meter. Keith Dienes of the University of Arizona, Emilian Dudas of the University of Paris–South and Tony Gherghetta, now at the University of Minnesota, observed in 1998 that extra dimensions smaller than 10^{-19} meter could allow the forces to unify at distances much larger than 10^{-32} meter.

Since our proposal in 1998 a number of interesting variations have appeared, using the same basic ingredients of extra dimensions and our universe-on-a-wall. In an intriguing model, Lisa Randall of Harvard University and Raman Sundrum of Johns Hopkins University proposed that gravity itself may be concentrated on a membrane in a five-dimensional spacetime that is infinite in all directions. Gravity appears very weak in our universe in a natural way if we are on a different membrane.

For 20 years, the conventional approach to tackling the hierarchy problem, and therefore understanding why gravity is so weak, had been to assume that the Planck scale near 10^{-35} meter is fundamental and that particle physics must change near 10^{-19} meter. Quantum gravity would remain in the realm of speculation, hopelessly out of the reach of experiment. We now realize this does not have to be the case. If there are large new dimensions, in the next several years we could discover deviations from Newton's law near 6×10^{-5} meter, say, and we would detect stringy vibrations or black holes at the LHC. Quantum gravity and string theory would become testable science. Whatever happens, experiment will point the way to answering a 300-year-old question. By 2010 we will have made decisive progress toward understanding why gravity is so weak. And we may find that we live in a strange Flatland, a membrane universe where quantum gravity is just around the corner. SA

MORE TO EXPLORE

The Theory Formerly Known as Strings. Michael Duff in *Scientific American,* Vol. 278, No. 2, pages 64–69; February 1998.

The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory. Brian Greene. W. W. Norton, 1999.

Flatland: A Romance of Many Dimensions. Edwin A. Abbott. Text is available from the Gutenberg project at **ftp://ibiblio.org/pub/docs/books/gutenberg/etext94/flat11.txt**

An introduction to tabletop gravity experiments is available at **http://mist.npl.washington.edu/eotwash/**

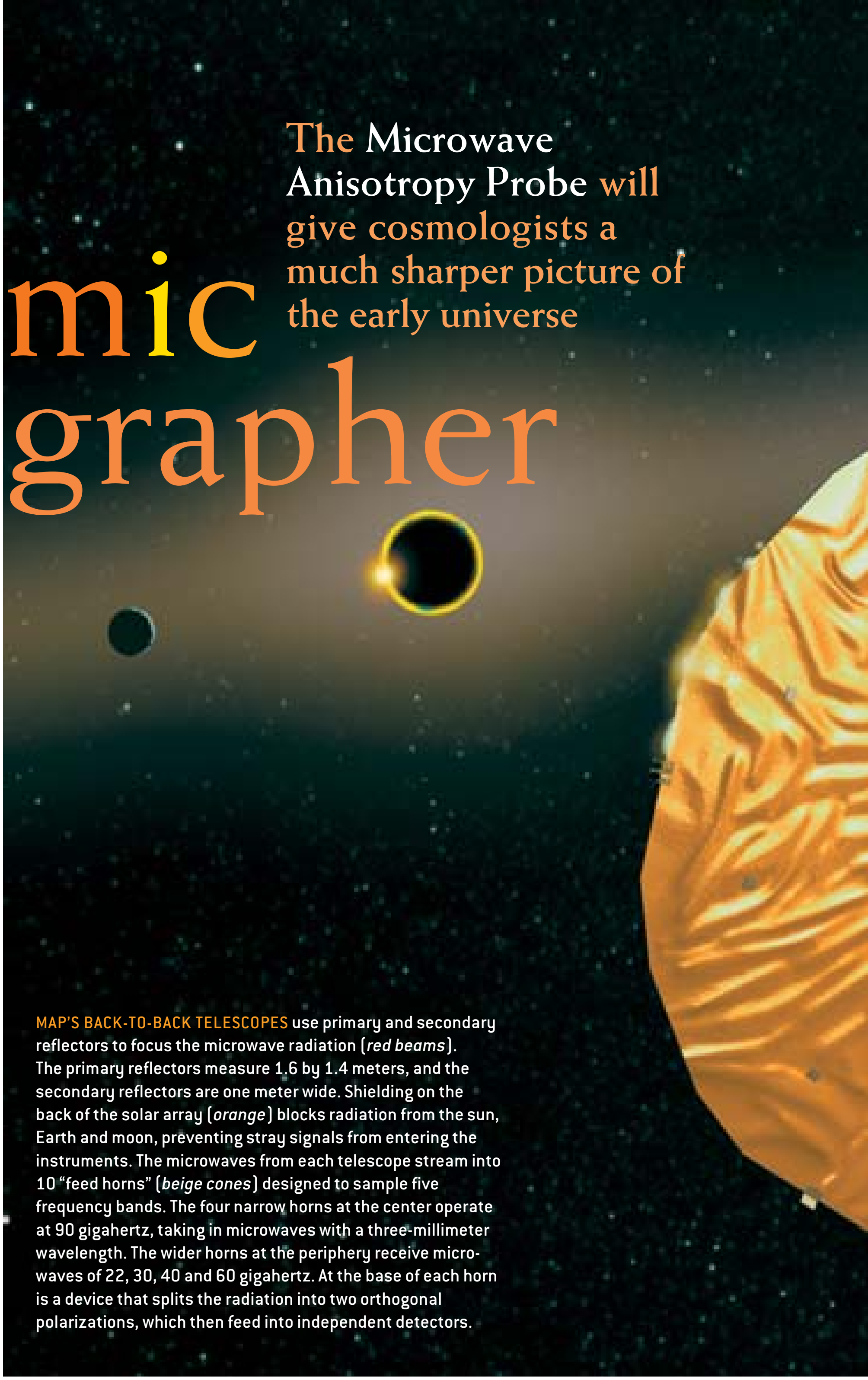
An introduction to string theory is available at **http://superstringtheory.com/**

DESTINY

a cosmic cartographer

The Microwave Anisotropy Probe will give cosmologists a much sharper picture of the early universe

BY CHARLES L. BENNETT, GARY F. HINSHAW AND LYMAN PAGE



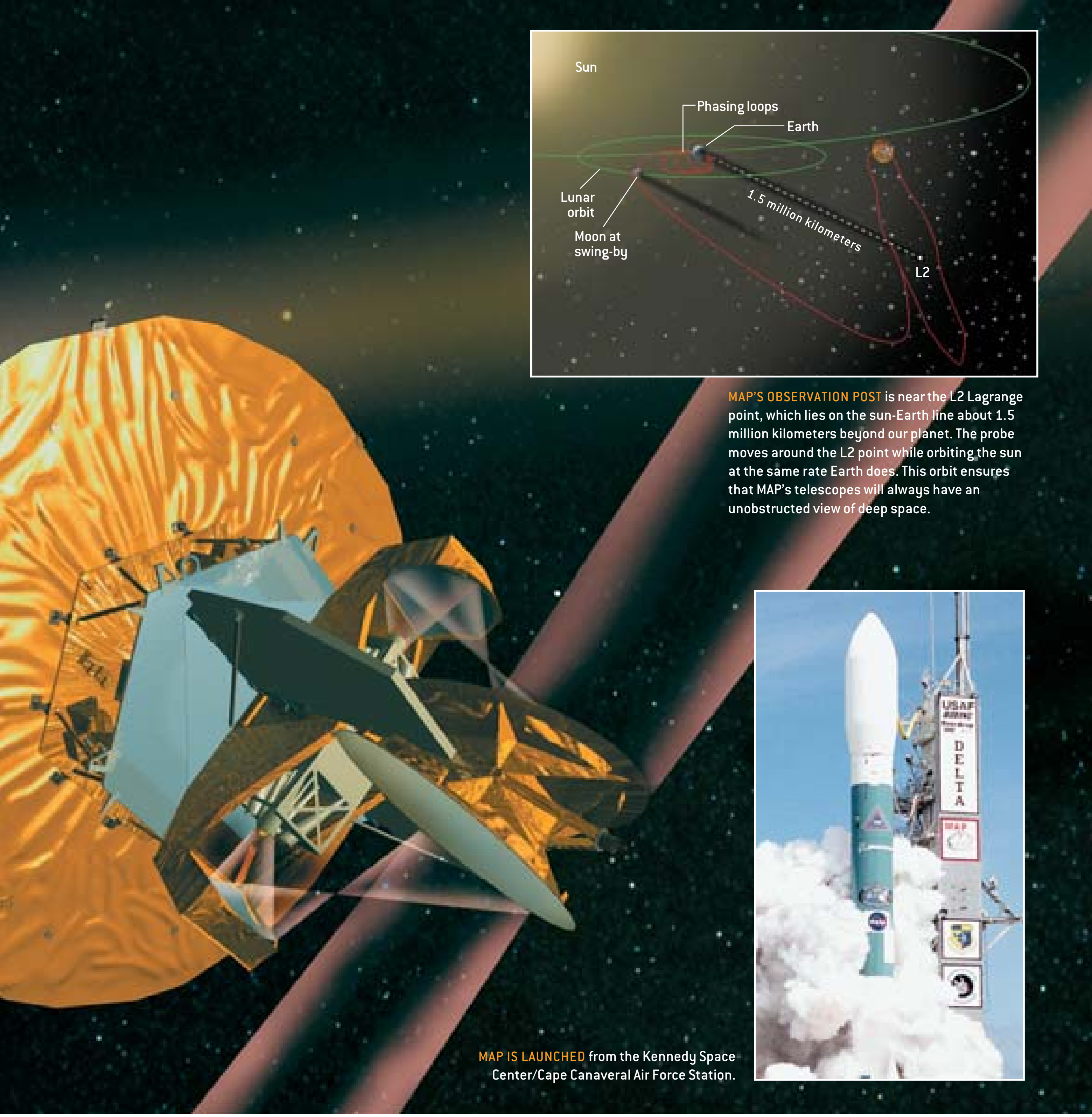
MAP'S BACK-TO-BACK TELESCOPES use primary and secondary reflectors to focus the microwave radiation (*red beams*). The primary reflectors measure 1.6 by 1.4 meters, and the secondary reflectors are one meter wide. Shielding on the back of the solar array (*orange*) blocks radiation from the sun, Earth and moon, preventing stray signals from entering the instruments. The microwaves from each telescope stream into 10 "feed horns" (*beige cones*) designed to sample five frequency bands. The four narrow horns at the center operate at 90 gigahertz, taking in microwaves with a three-millimeter wavelength. The wider horns at the periphery receive microwaves of 22, 30, 40 and 60 gigahertz. At the base of each horn is a device that splits the radiation into two orthogonal polarizations, which then feed into independent detectors.

On June 30, 2001, NASA launched a Delta 2 rocket carrying an 840-kilogram, four-meter-high spacecraft. Over the next three months the Microwave Anisotropy Probe (MAP) maneuvered into its orbit around the sun, 1.5 million kilometers beyond Earth's orbit. MAP is now observing the cosmic microwave background (CMB) radiation in exquisite detail over the entire sky. Because this radiation was emitted nearly 15 billion years ago and has not interacted significantly with anything since then, getting a clear picture of the CMB is equivalent to seeing a map of the early universe. By studying this map, scientists can learn the composition, geometry and history of the cosmos.

MAP is designed to measure the anisotropy of the CMB—the minuscule variations in the temperature of the radiation coming from different parts of the sky. MAP can record fluctuations as small as 20 millionths of a kelvin from the radiation's average temperature of 2.73 kelvins. What is more, the probe can detect hot and cold spots that subtend less than 0.23 degree across the sky, yielding a total of about one million measurements. Thus, MAP's observations of the CMB will be far more detailed than the previous full-sky map, produced in the early 1990s by the Cosmic Background Explorer (COBE), which was limited to a seven-degree angular resolution.

One reason for the improvement is that MAP employs two microwave telescopes, placed back-to-back, to focus the incoming radiation. The signals from the telescopes feed into 10 "differencing assemblies" that analyze five frequency bands in the CMB spectrum. But rather than measure the absolute temperature of the radiation, each assembly records the temperature difference between the signals from the two telescopes. Because the probe rotates, spinning once every two minutes and precessing once every hour, the differencing assemblies compare the temperature at each point in the sky with 1,000 other points, producing an interlocking set of data. The strategy is analogous

DON DIXON (*illustrations*); NASA (*photograph*)



MAP'S OBSERVATION POST is near the L2 Lagrange point, which lies on the sun-Earth line about 1.5 million kilometers beyond our planet. The probe moves around the L2 point while orbiting the sun at the same rate Earth does. This orbit ensures that MAP's telescopes will always have an unobstructed view of deep space.

MAP IS LAUNCHED from the Kennedy Space Center/Cape Canaveral Air Force Station.

to measuring the relative heights of bumps on a high plateau rather than recording each bump's elevation above sea level.

This method cancels out errors resulting from slight changes in the temperature of the spacecraft itself. The overall calibration of the data is done through a continuous measurement of the CMB dipole moment, the change in radiation temperature caused by Earth's motion through the cosmos. The guiding principle of MAP's design is to eliminate any spurious signals that might contaminate its measurements of the CMB. All indications are that MAP will begin returning high-quality results in January 2003, as scheduled. The cosmic map it will produce, of unprecedented fidelity, will be MAP's legacy for cosmology. SA

MAP SCIENCE TEAM includes Charles L. Bennett (NASA Goddard Space Flight Center), Mark Halpern (University of British Columbia), Gary F. Hinshaw (NASA GSFC), Norman C. Jarosik (Princeton University), Alan J. Kogut (NASA GSFC), Michele Limon (Princeton), Stephan S. Meyer (University of Chicago), Lyman Page (Princeton), David N. Spergel (Princeton), Gregory S. Tucker (Brown University), David T. Wilkinson (Princeton), Edward J. Wollack (NASA GSFC) and Edward L. Wright (University of California, Los Angeles).

Scientists may soon glimpse the universe's beginnings by studying the subtle ripples made by gravitational waves

Echoes from the big bang

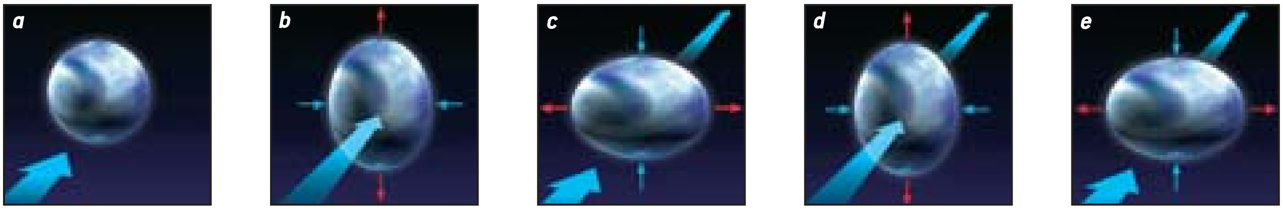
BY ROBERT R. CALDWELL AND MARC KAMIONKOWSKI

Cosmologists are still asking the same questions that the first stargazers posed as they surveyed the heavens. Where did the universe come from? What, if anything, preceded it? How did the universe arrive at its present state, and what will be its future? Although theorists have long speculated on the origin of the cosmos, until recently they had no way to probe the universe's earliest moments to test their hypotheses. In recent years, however, researchers have identified a method for observing the universe as it was in the very first fraction of a second after the big bang. This method involves looking for traces of gravitational waves in the cosmic microwave background (CMB), the cooled radiation that has permeated the universe for nearly 15 billion years.

The CMB was emitted about 400,000 years after the big bang, when electrons and protons in the primordial plasma—the hot, dense soup of subatomic particles that filled the early universe—first combined to form hydrogen atoms. Because this radiation provides a snapshot of the universe at that time, it has become the Rosetta stone of cosmology. After the CMB was discovered in 1965, researchers found that its temperature—a measure of the intensity of the black body radiation—was very close to 2.7 kelvins, no matter which direction they looked in the sky. In other words, the CMB appeared to be isotropic, which indicated that the early universe was remarkably uniform. In the early 1990s, though, a satellite called the Cosmic Background Explorer (COBE) detected minuscule variations—only one part in 100,000—in the radiation's temperature. These variations provide evidence of small lumps and bumps in the primordial plasma. The inhomogeneities in the distribution of mass later evolved into the large-scale structures of the cosmos: the galaxies and galaxy clusters that exist today.

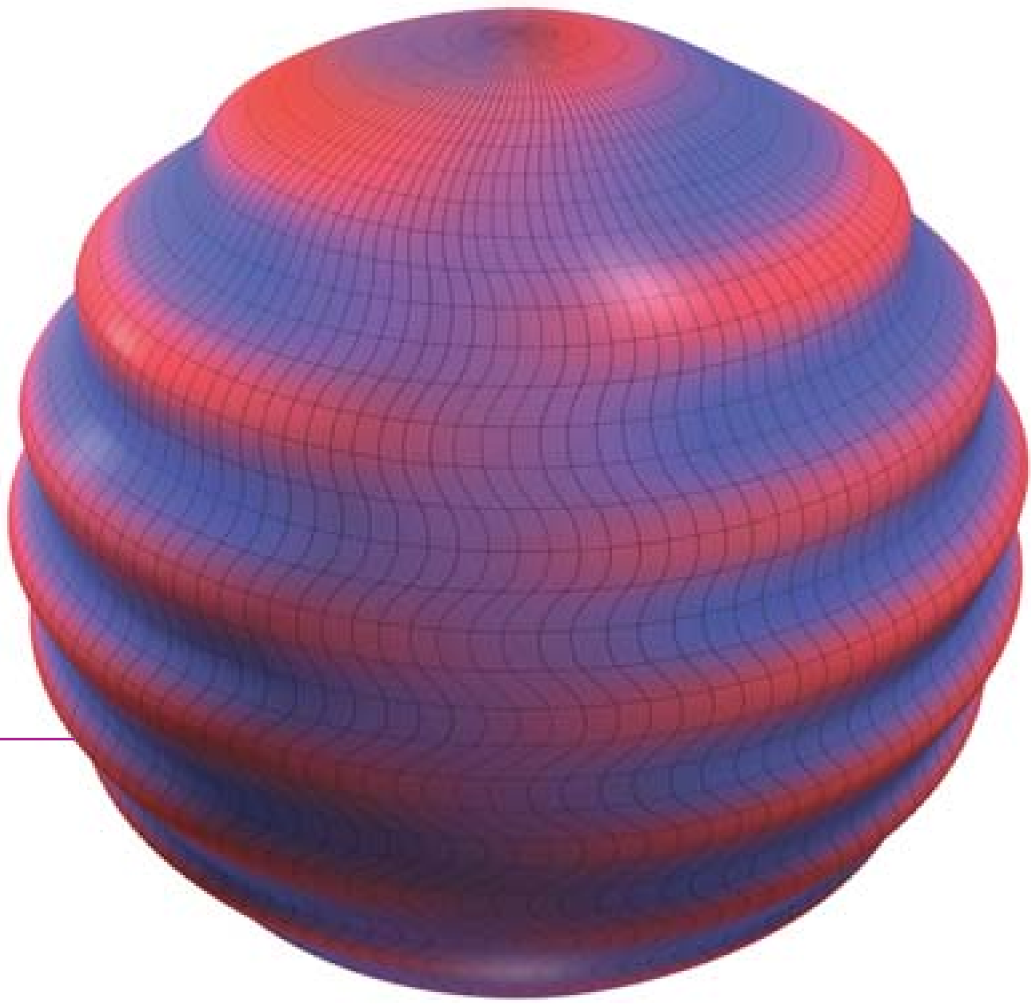
In the late 1990s several ground-based and balloon-borne detectors observed the CMB with much finer angular resolution than COBE did, revealing structures in the primordial plasma that subtend less than one degree across the sky. (For comparison, the moon subtends about half a degree.) The size of the primordial structures indicates that the geometry of the universe is flat. The observations are also consistent with the theory of inflation, which postulates that an epoch of phenomenally rapid cosmic expansion took place in the first few moments after the big bang. In June 2001 NASA launched the Microwave Anisotropy Probe (MAP), to extend the precise observations of the CMB to the entire sky [see "A Cosmic Cartographer," on page 74]. Currently gathering data from its deep orbit 1.5 million kilometers beyond the earth, MAP is expected to deliver its first scientific results by early 2003. The European Space Agency's Planck spacecraft, to launch in 2007, will conduct an even more detailed mapping. Cosmologists expect that the observations will unearth a treasure trove of information about the early universe.

In particular, researchers are hoping to find direct evidence of the epoch of inflation. The strongest evidence—the "smoking gun"—would be the observation of inflationary gravitation-



GRAVITATIONAL WAVES

Although gravitational waves have never been directly observed, theory predicts that they can be detected because they stretch and squeeze the space they travel through. On striking a spherical mass (*a*), a wave first stretches the mass in one direction and squeezes it in a perpendicular direction (*b*). Then the effects are reversed (*c*), and the distortions oscillate at the wave's frequency (*d* and *e*). The distortions shown here have been greatly exaggerated; gravitational waves are usually too weak to produce measurable effects.



DISTORTED UNIVERSE

The fantastically rapid expansion of the universe immediately after the big bang should have produced gravitational waves. These waves would have stretched and squeezed the primordial plasma, inducing motions in the spherical surface that emitted the cosmic microwave background, or CMB. These motions, in turn, would have caused redshifts and blueshifts in the radiation's temperature and polarized the CMB. The illustration above shows the effects of a gravitational wave traveling from pole to pole, with a wavelength that is one quarter the radius of the sphere.

ILLUSTRATIONS BY SLIM FILMS

al waves. In 1918 Albert Einstein predicted the existence of gravitational waves as a consequence of his theory of general relativity. They are analogues of electromagnetic waves, such as x-rays, radio waves and visible light, which are moving disturbances of an electromagnetic field. Gravitational waves are moving disturbances of a gravitational field. Like light or radio waves, gravitational waves can carry information and energy from the sources that produce them. Moreover, gravitational waves can travel unimpeded through material that absorbs all forms of electromagnetic radiation. Just as x-rays allow doctors to peer through substances that visible light cannot penetrate, gravitational waves should allow researchers to view astrophysical phenomena that cannot be seen otherwise. Although gravitational waves have never been directly detected, astronomical observations have confirmed that pairs of extremely dense objects such as neutron stars and black holes generate the waves as they spiral toward each other.

The plasma that filled the universe during its first 400,000 years was opaque to electromagnetic radiation, because any emitted photons were immediately scattered in the soup of subatomic particles. Therefore, astronomers cannot observe any electromagnetic signals dating from before the CMB. In contrast, gravitational waves could propagate through the plasma. What is more, the theory of inflation predicts that the explosive expansion of the universe 10^{-38} second after the big bang should have produced gravitational waves. If the theory is correct, these waves would have echoed across the early universe and, 400,000 years later, left subtle ripples in the CMB that can be observed today.

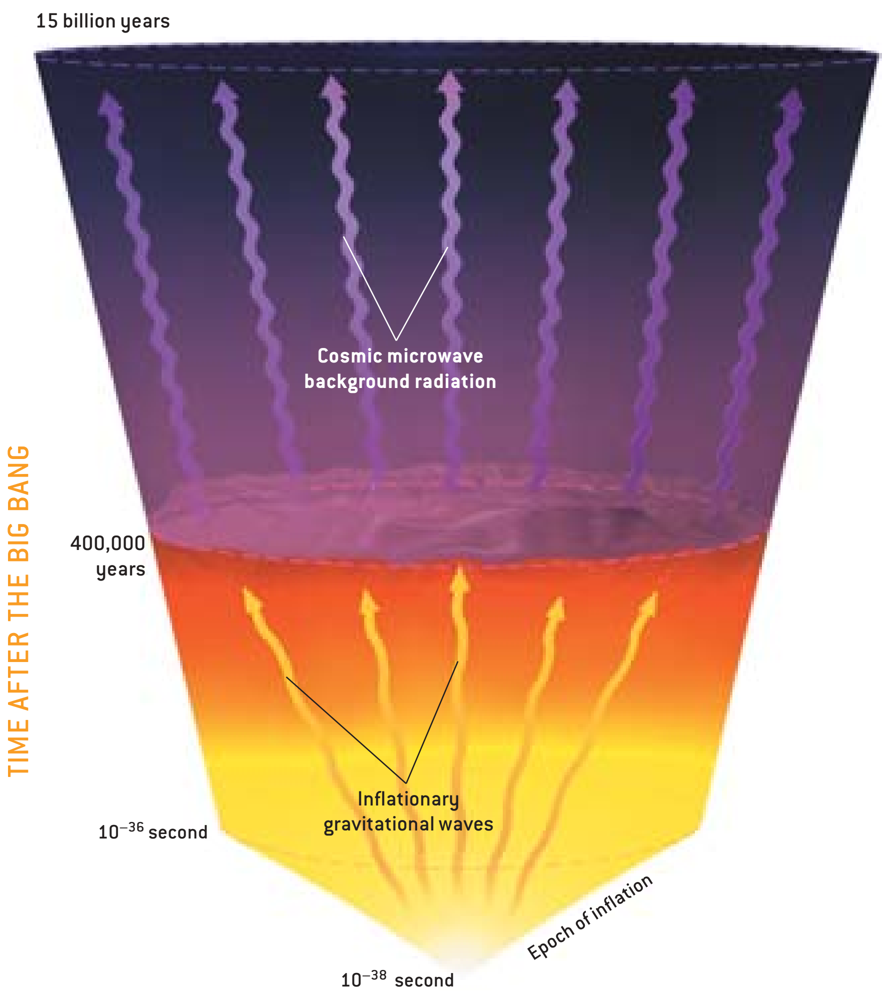
Waves from Inflation

TO UNDERSTAND HOW inflation could have produced gravitational waves, let's examine a fascinating consequence of quantum mechanics: empty space is not so empty. Virtual pairs of particles are spontaneously created and destroyed all the time. The Heisenberg uncertainty principle declares that a pair of particles with energy ΔE may pop into existence for a time Δt before they annihilate each other, provided that $\Delta E\ \Delta t < \hbar/2$, where $\hbar$ is the reduced Planck's constant (1.055×10^{-34} joule-second). You need not worry, though, about virtual apples or bananas popping out of empty space, because the formula applies only to elementary particles and not to complicated arrangements of atoms.

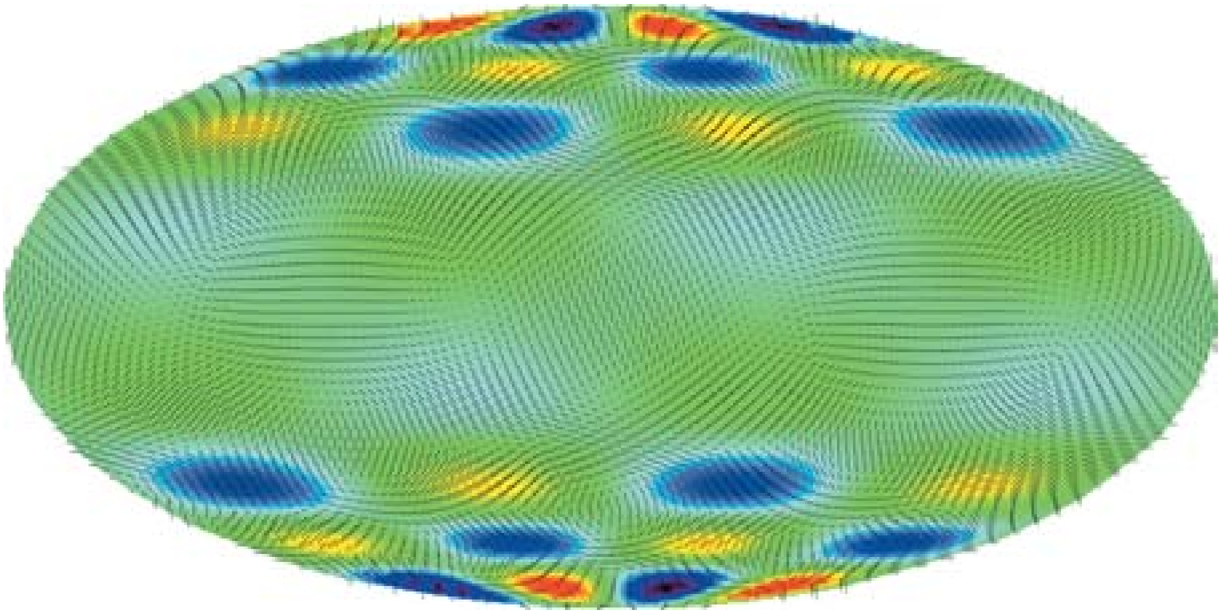
One of the elementary particles affected by this process is the graviton, the quantum particle of gravitational waves (analogous to the photon for electromagnetic waves). Pairs of virtual gravitons are constantly popping in and out of existence. During inflation, however, the virtual gravitons would have been pulled apart much faster than they could have disappeared back into the vacuum. In essence, the virtual particles would have become real particles. Furthermore, the fantastically rapid expansion of the universe would have stretched the graviton wavelengths from microscopic to macroscopic lengths. In this way, inflation would have pumped energy into the production of gravitons, generating a spectrum of gravitational waves that reflected the conditions in the universe in those first moments after the big bang. If inflationary gravitational waves do indeed exist, they would be the oldest relic in the universe, created 400,000 years *before* the CMB was emitted.

■ COSMIC TIME LINE

During the epoch of inflation—the tremendous expansion of the universe that took place in the first moments after the big bang—quantum processes generated a spectrum of gravitational waves. The waves echoed through the primordial plasma, distorting the CMB radiation that was emitted about 400,000 years later. By carefully observing the CMB today, cosmologists may detect the plasma motions induced by the inflationary waves.



ALFRED T. KAMAJIAN



RELIC IN THE RADIATION

Inflationary gravitational waves would have left a distinctive imprint on the CMB. This illustration depicts the simulated temperature variations and polarization patterns that would result from the distortions shown in the bottom illustration on page 77. The red and blue spots represent colder and hotter regions of the CMB, and the small line segments indicate the orientation angle of the polarization in each region of the sky.

COURTESY OF LEVEN WADLEY *Columbia University*

Whereas the microwave radiation in the CMB is largely confined to wavelengths between one and five millimeters (with a peak intensity at two millimeters), the wavelengths of the inflationary gravitational waves would span a much broader range: one centimeter to 10^{23} kilometers, which is the size of the present-day observable universe. The theory of inflation stipulates that the gravitational waves with the longest wavelengths would be the most intense and that their strength would depend on the rate at which the universe expanded during the inflationary epoch. This rate is proportional to the energy scale of inflation, which was determined by the temperature of the universe when inflation began. And because the universe was hotter at earlier times, the strength of the gravitational waves ultimately depends on the time at which inflation started.

Unfortunately, cosmologists cannot pinpoint this time, because they do not know in detail what caused inflation. Some physicists have theorized that inflation started when three of the fundamental interactions—the strong, weak and electromagnetic forces—became dissociated soon after the universe's creation. According to this theory, the three forces were one and the same at the very beginning but became distinct 10^{-38} second after the big bang, and this event somehow triggered the sudden expansion of the cosmos. If the theory is correct, inflation would have had an energy scale of 10^{15} to 10^{16} GeV. (One GeV is the energy a proton would acquire while being accelerated through a voltage drop of one billion volts. The largest particle accelerators currently reach energies of 10^3 GeV.) On the other hand, if inflation were triggered by another physical phenomenon occurring at a later time, the gravitational waves would be weaker.

Once produced during the first fraction of a second after the big bang, the inflationary gravitational waves would propagate forever, so they should still be running across the universe. But how can cosmologists observe them? First consider how an ordinary stereo receiver detects a radio signal. The radio waves consist of oscillating electrical and magnetic fields, which cause the electrons in the receiver's antenna to move back and forth. The motions of these electrons produce an electric current that the receiver records.

Similarly, a gravitational wave can induce an oscillatory stretching and squeezing of the space it travels through. These oscillations would cause small motions in a set of freely floating test masses. In the late 1950s physicist Hermann Bondi of King's College London tried to convince skeptics of the physical reality of such waves by describing a hypothetical gravitational-wave detector. The idealized apparatus was a pair of rings hanging freely on a long, rigid bar. An incoming gravitational wave of amplitude h and frequency f would cause the distance L between the two rings to alternately contract and expand by an amount $h \times L$, with a frequency f. The heat from the friction of the rings rubbing against the bar would provide evidence that the gravitational wave carries energy.

Researchers are currently building sophisticated gravitational-wave detectors, which will use lasers to track the tiny motions of suspended masses [*see box on next page*]. The distance between the test masses determines the band of wave-

THE AUTHORS

ROBERT R. CALDWELL and *MARC KAMIONKOWSKI* were both physics majors in the class of 1987 at Washington University. Caldwell earned his Ph.D. in physics at the University of Wisconsin–Milwaukee in 1992. One of the chief formulators of the theory of quintessence, Caldwell is now assistant professor of physics and astronomy at Dartmouth College. Kamionkowski earned his doctorate in physics at the University of Chicago in 1991. Now professor of theoretical physics and astrophysics at the California Institute of Technology, he received the Warner Prize in 1998 for his contributions to theoretical astronomy.

Wave Hunters

New detectors will soon be ready

THE GRAVITATIONAL WAVES produced by quantum processes during the inflationary epoch are by no means the only ones believed to be traveling across the universe. Many astrophysical systems, such as orbiting binary stars, merging neutron stars and colliding black holes, should also emit powerful gravitational waves.

The problem with detecting the waves is that their strength fades as they spread outward while traveling hundreds of millions of light-years to the earth. To measure such minuscule oscillations, researchers are preparing the Laser Interferometer Gravitational-Wave Observatory (LIGO), which consists of facilities in Livingston, La. (*above*), and Hanford, Wash. Results from the facilities will be compared to rule out local effects that mimic gravitational waves, such as seismic activity, acoustic noise and laser instabilities [see "Ripples in Spacetime," on page 88].

Physicists are also building smaller detectors that will work in tandem with LIGO, allowing researchers to triangulate the sources of gravitational waves. Examples of these observatories are TAMA (near Tokyo), Virgo (near Pisa, Italy) and GEO (near Hannover, Germany). And to monitor gravitational waves with longer wavelengths, NASA and the European Space Agency are planning to launch the Laser Interferometer Space Antenna in 2010. Unfortunately, none of these proposed observatories will be sensitive enough to detect the gravitational waves produced by inflation. Only the cosmic microwave background radiation can reveal their presence. *—R.R.C. and M.K.*

lengths that the devices can monitor. The largest of the ground-based detectors, which has a separation of four kilometers between the masses, will be able to measure the oscillations caused by gravitational waves with wavelengths from 30 to 30,000 kilometers; a planned space-based observatory may be able to detect wavelengths about 1,000 times longer. The gravitational waves generated by neutron star mergers and black hole collisions have wavelengths in this range, so they can be detected by the new instruments. But the inflationary gravitational waves in this range are much too weak to produce measurable oscillations in the detectors.

The strongest inflationary gravitational waves are those with the longest wavelengths, comparable to the diameter of the observable universe. To detect these waves, researchers need to observe a set of freely floating test masses separated by similarly large distances. Serendipitously, nature has provided just such an arrangement: the primordial plasma that emitted the CMB radiation. During the 400,000 years between the epoch of inflation and the emission of the CMB, the ultralong-wavelength gravitational waves echoed across the early universe, alternately stretching and squeezing the plasma [*see illustration on page 78*]. Researchers can observe these oscillatory motions today by looking for slight Doppler shifts in the CMB.

If, at the time when the CMB was emitted, a gravitational wave was stretching a region of plasma toward us—that is, toward the part of the universe that would eventually become our galaxy—the radiation from that region will appear bluer to observers because it has shifted to shorter wavelengths (and hence a higher temperature). Conversely, if a gravitational wave was squeezing a region of plasma away from us when the CMB was emitted, the radiation will appear redder because it has shifted to longer wavelengths (and a lower temperature). By surveying the blue and red spots in the CMB—which correspond to hotter and colder radiation temperatures—researchers could conceivably see the pattern of plasma motions induced by the inflationary gravitational waves. The universe itself becomes a gravitational-wave detector.

Particulars of Polarization

THE TASK IS NOT so simple, however. As we noted at the beginning of this article, mass inhomogeneities in the early universe also produced temperature variations in the CMB. (For example, the gravitational field of the denser regions of plasma would have redshifted the photons emitted from those regions, producing some of the temperature differences observed by COBE.) If cosmologists look at the radiation temperature alone, they cannot tell what fraction (if any) of the variations should be attributed to gravitational waves. Even so, scientists at least know that gravitational waves could not have produced any more than the one-in-100,000 temperature differences observed by COBE and the other CMB radiation detectors. This fact puts an interesting constraint on the physical phenomena that gave rise to inflation: the energy scale of inflation must be less than about 10^{16} GeV, and therefore the epoch could not have occurred earlier than 10^{-38} second after the big bang.

But how can cosmologists go further? How can they get around the uncertainty over the origin of the temperature fluctuations? The answer lies with the *polarization* of the CMB. When light strikes a surface in such a way that the light scatters at nearly a right angle from the original beam, it becomes linearly polarized—that is, the waves become oriented in a particular direction. This is the effect that polarized sunglasses exploit: because the sunlight that scatters off the ground is typically polarized in a horizontal direction, the filters in the glasses reduce the

ALEX DESSELLE *Skyview Technologies*

glare by blocking light waves with this orientation. The CMB is polarized as well. Just before the early universe became transparent to radiation, the CMB photons scattered off the electrons in the plasma for the last time. Some of these photons struck the particles at large angles, which polarized the radiation.

The key to detecting the inflationary gravitational waves is the fact that the plasma motions caused by the waves produced a different pattern of polarization than the mass inhomogeneities did. The idea is relatively simple. The linear polarization of the CMB can be depicted with small line segments that show the orientation angle of the polarization in each region of the sky [*see illustration on page 79*]. These line segments are sometimes arranged in rings or in radial patterns. The segments can also appear in rotating swirls that are either right- or left-handed—that is, they seem to be turning clockwise or counterclockwise [*see illustration at right*].

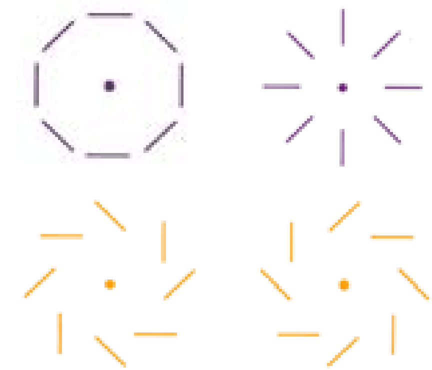
The "handedness" of these patterns is the clue to their origin. The mass inhomogeneities in the primordial plasma could not have produced such polarization patterns, because the dense and rarefied regions of plasma had no right- or left-handed orientation. In contrast, gravitational waves do have a handedness: they propagate with either a right- or left-handed screw motion. The polarization pattern produced by gravitational waves will look like a random superposition of many rotating swirls of various sizes. Researchers describe these patterns as having a curl, whereas the ringlike and radial patterns produced by mass inhomogeneities have no curl.

Not even the most keen-eyed observer can look at a polarization diagram, such as the one shown on page 79, and tell by eye whether it contains any patterns with curls. But an extension of Fourier analysis—a mathematical technique that can break up an image into a series of waveforms—can be used to divide a polarization pattern into its constituent curl and curl-free patterns. Thus, if cosmologists can measure the CMB polarization and determine what fraction came from curl patterns, they can calculate the amplitude of the ultralong-wavelength inflationary gravitational waves. Because the amplitude of the waves was determined by the energy of inflation, researchers will get a direct measurement of that energy scale. This finding, in turn, will help answer the question of whether inflation was triggered by the unification of fundamental forces.

What are the prospects for detecting the gentle rings and curls and swirls of the polarized CMB sky? That is the next goal of CMB scientists. Although theorists are confident that the relic radiation is polarized, observational verification has eluded researchers. Roughly a dozen experiments worldwide are striving to measure the variation of the polarization pattern, and we can expect exciting results soon. But theorists also predict that the strength of the curl-free polarization component is much stronger than the curl component—the "smoking gun" of the inflationary gravitational waves we have described. So though there is a good chance the MAP satellite or one of the ground- or balloon-based experiments will detect the CMB's curl-free polarization within the next year, the curl component will remain just out of reach.

Subsequent experiments may have a better chance. If inflation was indeed caused by the unification of forces, its gravitational-wave signal may be strong enough to be detected by the Planck satellite, although an even more sensitive next-generation spacecraft might be needed. CMB scientists are already working with NASA to plan such a mission: the Cosmic Microwave Background Polarization Experiment (CMBPOL), which would fly sometime after 2014. If the inflationary theory is true, and there are ultralong-wavelength gravitational waves of primordial origin coursing through the cosmos, then CMBPOL will be able to sense the telltale signs of the squeezing and stretching of the plasma at last scattering. The discovery would extend our understanding of the universe back to the earliest fraction of a second after the big bang. But if inflation was triggered by other physical phenomena occurring at later times and lower energies, the signal from the gravitational waves will be far too weak to be detected in the foreseeable future.

Because cosmologists are not certain about the origin of inflation, they cannot definitively predict the strength of the polarization signal produced by inflationary gravitational waves. But if there is even a small chance that the signal is detectable, then it is worth pursuing. Its detection would not only provide incontrovertible evidence of inflation but also give us the extraordinary opportunity to look back at the very earliest times, just 10^{-38} second after the big bang. We could then contemplate addressing one of the most compelling questions of the ages: Where did the universe come from? SA



POLARIZATION PATTERNS

The polarization of the CMB may hold important clues to the history of the early universe. Density variations in the primordial plasma would cause ringlike and radial patterns of polarization (*top*). Gravitational waves, in contrast, would produce right- and left-handed swirls (*bottom*).

COURTESY OF ROBERT R. CALDWELL AND MARC KAMIONKOWSKI

MORE TO EXPLORE

First Space-Based Gravitational-Wave Detectors. Robert R. Caldwell, Marc Kamionkowski and Leven Wadley in *Physical Review D,* Vol. 59, Issue 2, pages 27101–27300; January 15, 1999.

Recent observations of the cosmic microwave background are described at these Web sites: **pupgg.princeton.edu/~cmb/**; **www.physics.ucsb.edu/~boomerang/**; **cosmology.berkeley.edu/group/cmb/**

Details of the MAP and Planck missions are available at **map.gsfc.nasa.gov/**; **astro.estec.esa.nl/astrogen/planck/mission_top.html**

More information on gravitational-wave detectors is available at **www.ligo.caltech.edu**; **lisa.jpl.nasa.gov**

DESTINY

Exploring Our Universe and Others

IN THIS CENTURY COSMOLOGISTS WILL UNRAVEL THE MYSTERY OF OUR UNIVERSE'S BIRTH— AND PERHAPS PROVE THE EXISTENCE OF OTHER UNIVERSES AS WELL

BY MARTIN REES

Cosmic exploration is preeminently a 20th-century achievement. Only in the 1920s did we realize that our Milky Way, with its 100 billion stars, is just one galaxy among millions. Our empirical knowledge of the universe has been accumulating ever since. We can now set our entire solar system in a grand evolutionary context, tracing its constituent atoms back to the initial instants of the big bang. If we were ever to discover alien intelligences, one thing we might share with them—perhaps the only thing—would be a common interest in the cosmos from which we have all emerged.

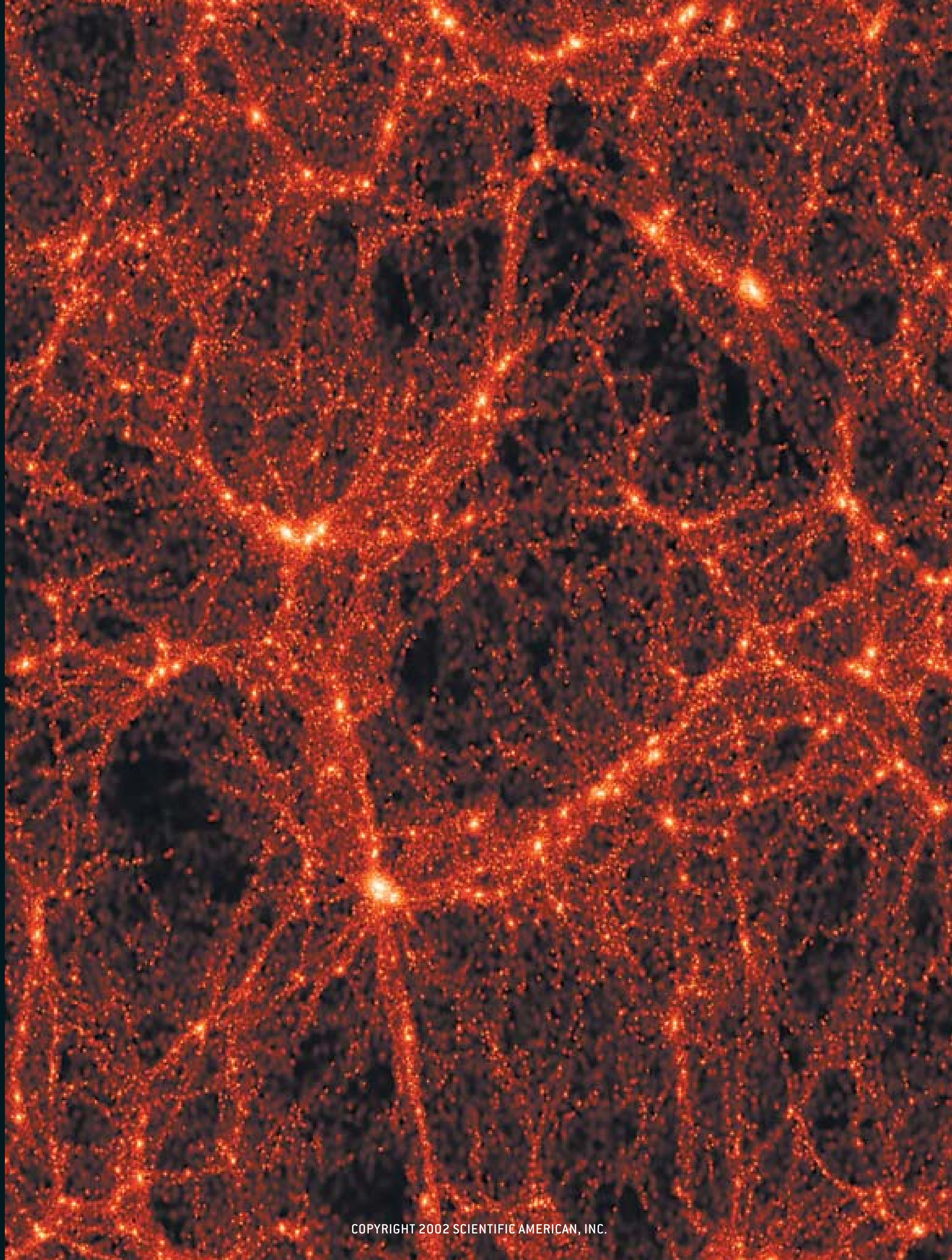
Using the current generation of ground-based and orbital observatories, astronomers can look back into the past and see plain evidence of the evolution of the universe. Marvelous images from the Hubble Space Telescope reveal galaxies as they were in remote times: balls of glowing, diffuse gas dotted with massive, fast-burning blue stars. These stars transmuted the pristine hydrogen from the big bang into heavier atoms, and when the stars died they seeded their galaxies with the basic building blocks of planets and life—carbon, oxygen, iron and so on. A Creator didn't have to turn 92 different knobs to make all the naturally occurring elements in the periodic table. Instead the galaxies act as immense ecosystems, forging elements and recycling gas through successive generations of stars. The human race itself is composed of stardust—or, less romantically, the nuclear waste from the fuel that makes stars shine.

Astronomers have also learned much about the earlier, pregalactic era by studying the microwave background radiation that makes even intergalactic space slightly warm. This afterglow of creation tells us that the entire universe was once hotter than the centers of stars. Scientists can use laboratory data to calculate how much nuclear fusion would have happened during the first few minutes after the big bang. The predicted proportions of hydrogen, deuterium and helium accord well with what astronomers have observed, thereby corroborating the big bang theory.

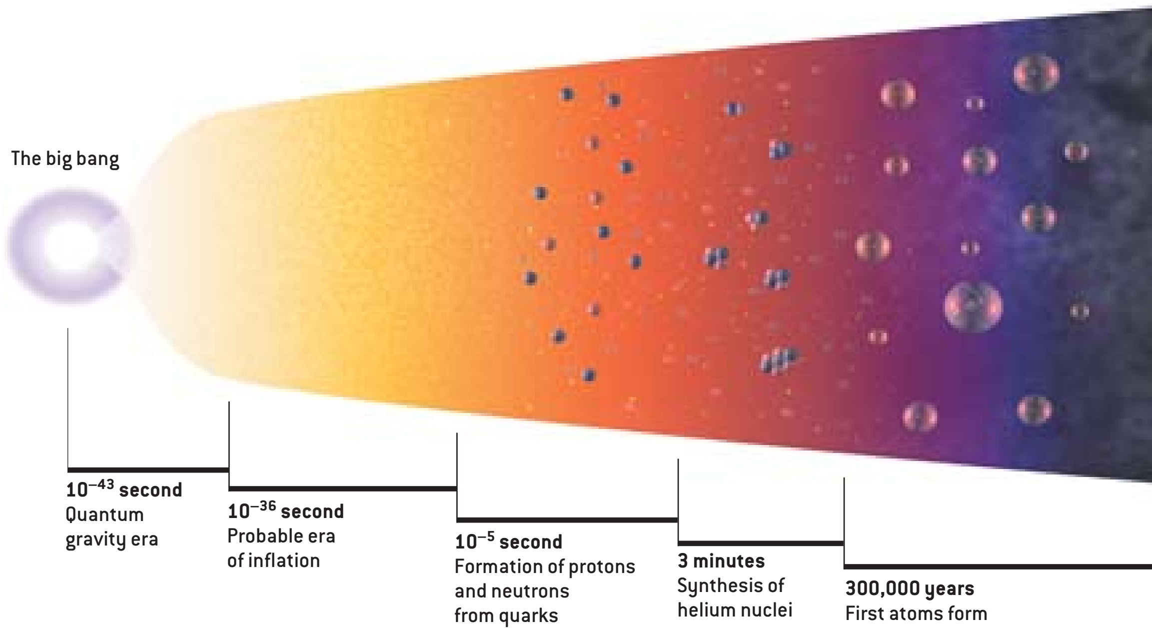
At first sight, attempts to fathom the cosmos might seem

LARGE-SCALE STRUCTURE of the universe can be simulated by running cosmological models on a supercomputer. In this simulation, produced by the Virgo Consortium, each particle represents a galaxy.

VIRGO CONSORTIUM (J. M. COLBERG, H.M.P. COUCHMAN, G. EFSTATHIOU, C. S. FRENK, A. JENKINS, J. A. PEACOCK, F. R. PEARCE AND S.D.M. WHITE)



COSMIC TIME LINE shows the evolution of our universe from the big bang to the present day. In the first instant of creation—the epoch of inflation—the universe expanded at a staggering rate. After about three minutes, the plasma of particles and radiation cooled enough to allow the formation of simple atomic nuclei; after another 300,000 years, atoms of hydrogen and helium began to form. The first stars and galaxies appeared about a billion years later. The ultimate fate of the universe—whether it will expand forever or recollapse—is still unknown, although current evidence favors perpetual expansion.



presumptuous and premature, even at the start of the 21st century. Cosmologists have, nonetheless, made remarkable progress in recent years. This is because what makes things baffling is their degree of complexity, not their sheer size—and a star is simpler than an insect. The fierce heat within stars, and in the early universe, guarantees that everything breaks down into its simplest constituents. It is the biologists, whose role it is to study the intricate multilayered structure of trees, butterflies and brains, who face the tougher challenge.

The progress in cosmology has brought new mysteries into sharper focus and raised questions that will challenge astronomers well into this century. For example, why does our universe contain its observed mix of ingredients? And how, from its dense beginnings, did it heave itself up to such a vast size? The answers will take us beyond the physics with which we are familiar and will require new insights into the nature of space and time. To truly understand the history of the universe, scientists must discover the profound links between the cosmic realm of the very large and the quantum world of the very small.

It is embarrassing to admit, but astronomers still don't know what our universe is made of. The objects that emit radiation that we can observe—such as stars, quasars and galaxies—constitute only a small fraction of the universe's matter. The vast bulk of matter is dark and unaccounted for. Most cosmologists believe dark matter is composed of weakly interacting particles left over from the big bang, but it could be something even more exotic. Whatever the case, it is clear that galaxies, stars and planets are a mere afterthought in a cosmos dominated by quite different stuff. Searches for dark matter, mainly via sensitive underground experiments designed to detect elusive subatomic particles, will continue apace in this decade. The stakes are high: success would not only tell us what most of the universe is made of but would also probably reveal some fundamentally new kinds of particles.

The ultimate fate of our universe—whether it continues expanding indefinitely or eventually changes course and collapses to the so-called big crunch—depends on the total amount of dark matter and the gravity it exerts. Current data indicate that the universe contains only about 30 percent of the matter that would be needed to halt the expansion. (In cosmologists' jargon, omega—the ratio of observed density to critical density—is 0.3.) The odds favoring perpetual growth have recently strengthened further: tantalizing observations of distant supernovae suggest that the expansion of the universe may be speeding up rather than slowing down. These observations may indicate that an extra force overwhelms gravity on cosmic scales—a phenomenon perhaps related to what Albert Einstein called the cosmological constant, a form of energy latent in empty space itself that (unlike ordinary matter) has negative pressure and causes a repulsion. Studies of small nonuniformities in the background radiation reveal that our universe is "flat"—in the sense that the angles of a large triangle drawn in space add up to 180 degrees. Taken in conjunction with one another, these lines of evidence suggest that 5 percent of our universe (or slightly less) is composed of ordinary atoms, about 25 percent is dark matter and the other 70 percent is the even more perplexing dark energy.

Research is also likely to focus on the evolution of the universe's large-scale structure. If one had to answer the question

The great mystery for cosmologists is the series of events that occurred less than one millisecond after the big bang.

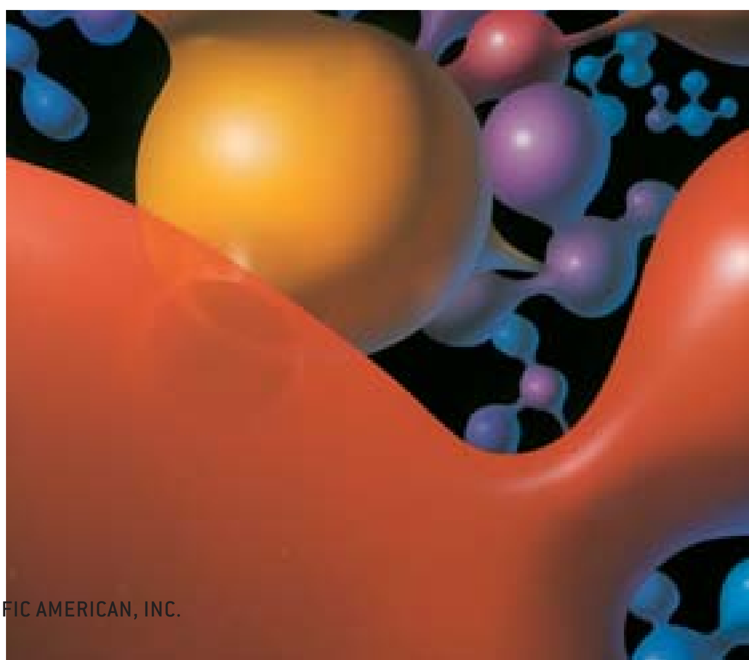
"What's been happening since the big bang?" in just one sentence, the best response might be to take a deep breath and say, "Ever since the beginning, gravity has been amplifying inhomogeneities, building up structures and enhancing temperature contrasts—a prerequisite for the emergence of the complexity that lies around us now and of which we're a part." Astronomers are now learning more about this 10-billion-year process by creating virtual universes on their computers. In the coming years, they will be able to simulate the history of the universe with ever improving realism and then compare the results with what telescopes reveal.

Questions of structure have preoccupied astronomers since the time of Isaac Newton, who wondered why all the planets circled the sun in the same direction and in almost the same plane. In his 1704 work *Opticks* he wrote: "Blind fate could never make all the planets move one and the same way in orbits concentrick." Such a wonderful uniformity in the planetary system, Newton believed, must be the effect of divine providence.

Now astronomers know that the coplanarity of the planets is a natural outcome of the solar system's origin as a spinning disk of gas and dust. Indeed, we have extended the frontiers of our knowledge to far earlier times; cosmologists can roughly outline the history of the universe back to the first second after the big bang. Conceptually, however, we're in little better shape than Newton was. Our understanding of the causal chain of events now stretches further back in time, but we still run into a barrier, just as surely as Newton did. The great mystery for cosmologists is the series of events that occurred less than one millisecond after the big bang, when the universe was extraordinarily small, hot and dense. The laws of physics with which we are familiar offer little firm guidance for explaining what happened during this critical period.

To unravel this mystery, cosmologists must first pin down—by improving and refining current observations—some of the characteristics of the universe when it was only one second old: its expansion rate, the size of its density fluctuations, and its proportions of ordinary atoms, dark matter and radiation. But to comprehend why our universe was set up this way, we must probe further back, to the first tiny fraction of a microsecond. Such an effort will require theoretical advances. Physicists must discover a way to relate Einstein's theory of general relativity, which governs large-scale interactions in the cosmos, to the quantum principles that apply at very short distances. A unified

MULTIPLE UNIVERSES are continually being born, according to some cosmologists. Each universe is shown here as an expanding bubble branching off from its parent universe. The changes in color represent shifts in the laws of physics from one universe to another.



ILLUSTRATIONS BY ALFRED T. KAMAJIAN

PAT RAWLINGS

LUNAR OBSERVATORIES will greatly extend the reach of 21st-century astronomers. The far side of the moon is an ideal place for telescopes because of its absence of atmosphere and its utterly dark nights, free of reflected sun and radio transmissions from Earth. Lunar ores can be used to build the instruments.

theory would be needed to explain what happened in the first crucial moments after the big bang, when the entire universe was squeezed into a space smaller than a single atom.

Astronomy is a subject in which observation is king. Now the same is true for cosmology—in contrast with the pre-1965 era, when speculation was largely unconstrained. The answers to many of cosmology's long-standing questions are most likely to come from the telescopes that have recently gone into use. The two Keck Telescopes on Mauna Kea in Hawaii are far more sensitive than earlier observatories and thus can glimpse fainter objects. Still more impressive is the European Southern Observatory's Very Large Telescope at Paranal in northern Chile—a linked array of four telescopes, each with mirrors eight meters in diameter. In space, astronomers can take advantage of the Chandra X-ray Observatory and its European counterpart, XMM-Newton. Several new instruments are under construction to detect radio waves, infrared emissions and cosmic rays. And a decade from now next-generation space telescopes will carry the enterprise far beyond what the Hubble can achieve.

Well before 2050 we will probably see the construction of giant observatories in space or perhaps on the far side of the moon. The sensitivity and imaging power of these arrays will vastly surpass that of any instruments now in use. The new telescopes will target black holes and planets in other solar systems. They will also provide snapshots of every cosmological era going back to the very first light, when the earliest stars (or maybe quasars) condensed out of the expanding debris from the big bang. Some of these observatories may even be able to measure gravitational waves, allowing scientists to probe vibrations in the fabric of spacetime itself.

The amount of data provided by all these instruments will be so colossal that the entire process of analysis and discovery will most likely be automated. Astronomers will focus their attention on heavily processed statistics for each population of objects they are studying and in this way find the best examples—for instance, the planets in other solar systems that are most like Earth. Researchers will also concentrate on extreme objects that may hold clues to physical processes that are not yet fully understood. One such object is the gamma-ray burster, which emits, for a few seconds, as much power as a billion galaxies. Increasingly, astronomers will use the heavens as a cosmic laboratory to probe phenomena that cannot be simulated on Earth.

Another benefit of automation will be open access to astronomical data that in the past were available to only a privileged few. Detailed maps of the sky will be available to anyone who can access or download them. Enthusiasts anywhere in the world will be able to check their own hunches, seek new patterns and discover unusual objects.

THE AUTHOR

MARTIN REES is Royal Society Research Professor at the University of Cambridge and holds the honorary title of Astronomer Royal. He was previously director of the Cambridge Institute of Astronomy and before that a professor at the University of Sussex. His research interests include black holes, galaxy formation and high-energy astrophysics. He also lectures and writes extensively for general audiences. His recent books are *Before the Beginning* (Perseus Publishing, 1998), *Just Six Numbers* (Basic Books, 2000) and *Our Cosmic Habitat* (Princeton University Press, 2001).

Intimations of a Multiverse?

COSMOLOGISTS VIEW the universe as an intricate tapestry that has evolved from initial conditions that were imprinted in the first microsecond after the big bang. Complex structures and phenomena have unfolded from simple physical laws—we wouldn't be here if they hadn't. Simple laws, however, do not necessarily lead to complex consequences. Consider an analogue from the field of fractal mathematics: the Mandelbrot set, a pattern with an infinite depth of structure, is encoded by a short algorithm, but other simple algorithms that are superficially similar yield very boring patterns.

Our universe could not have become structured if it were not expanding at a special rate. If the big bang had produced fewer density fluctuations, the universe would have remained dark, with no galaxies or stars. And there are other prerequisites for complexity. If our universe had more than three spatial dimensions, planets could not stay in orbits around stars. If gravity were much stronger, it would crush living organisms of human size, and stars would be small and short-lived. If nuclear forces were a few percent weaker, only hydrogen would be stable: there would be no periodic table, no chemistry and no life.

Some would argue that this fine-tuning of the universe, which seems so providential, is nothing to be surprised about, because we could not exist otherwise. There is, however, another interpretation: many universes may exist, but only some would allow creatures like us to emerge.

Perhaps, then, our big bang wasn't the only one. This speculation dramatically enlarges our concept of reality. The entire history of our universe becomes just an episode, a single facet, of the infinite multiverse. Some universes might resemble ours, but most would recollapse after a brief existence, or the laws governing them would not permit complex consequences.

Some cosmologists, especially Andrei Linde of Stanford University and Alex Vilenkin of Tufts University, have already shown how certain mathematical assumptions lead to the creation of a multiverse. But such ideas will remain on the speculative fringe of cosmology until we really understand—rather than just guess at—the extreme physics that prevailed immediately after the big bang. Will the long-awaited unified theory uniquely determine the masses of particles and the strengths of the basic forces? Or are these properties accidental outcomes of how our universe cooled—secondary manifestations of still deeper laws governing an entire ensemble of universes?

This topic might seem arcane, but the status of multiverse ideas affects how we should place our bets in some ongoing cosmological controversies. Some theorists have a strong preference for the simplest picture of the cosmos, which would require an omega of 1—the universe would be just dense enough to halt its own expansion. They are unhappy with observations suggesting that the universe is not nearly so dense and with extra complications such as the cosmological constant and dark energy. Perhaps we should draw a lesson from 17th-century astronomers Johannes Kepler and Galileo Galilei, who were upset to find that planetary orbits were elliptical. Circles, they thought, were simpler and more beautiful. But Newton later explained all orbits in terms of a simple, universal law of gravity. Had Galileo still been alive, he surely would have been joyfully reconciled to ellipses.

The parallel is obvious. If a low-density universe with a cosmological constant seems ugly, maybe this shows our limited vision. Just as Earth follows one of the few Keplerian orbits around the sun that allow it to be habitable, our universe may be one of the few habitable members of a grander ensemble.

Perhaps our big bang wasn't the only one; many universes may exist in the infinite multiverse.

The Next Challenges

SCIENTISTS ARE EXPANDING humanity's store of knowledge on three great frontiers: the very big, the very small and the very complex. Cosmology involves them all. In the coming years, researchers will focus on pinning down the basic universal constants, such as omega, and on discovering what dark matter is. I think there is a good chance of achieving both goals within 10 years. Maybe everything will fit the standard theoretical framework, and we will successfully determine not only the relative abundance of ordinary atoms and dark matter in the universe but also the cosmological constant and the primordial density fluctuations. If that happens, we will have taken the measure of our universe just as we have learned the size and shape of Earth and our sun. On the other hand, our universe may turn out to be too complicated to fit the standard framework. Some may describe the first outcome as optimistic; others may prefer to inhabit a more complicated and challenging place!

In addition, theorists must elucidate the exotic physics of the very earliest moments of the universe. If they succeed, we will learn whether there are many universes and which features of our universe are mere contingencies rather than the necessary outcomes of the deepest laws. Our understanding will still have limits, however. Physicists may someday discover a unified theory that governs all of physical reality, but they will never be able to tell us what breathes fire into their equations and what actualizes them in a real cosmos.

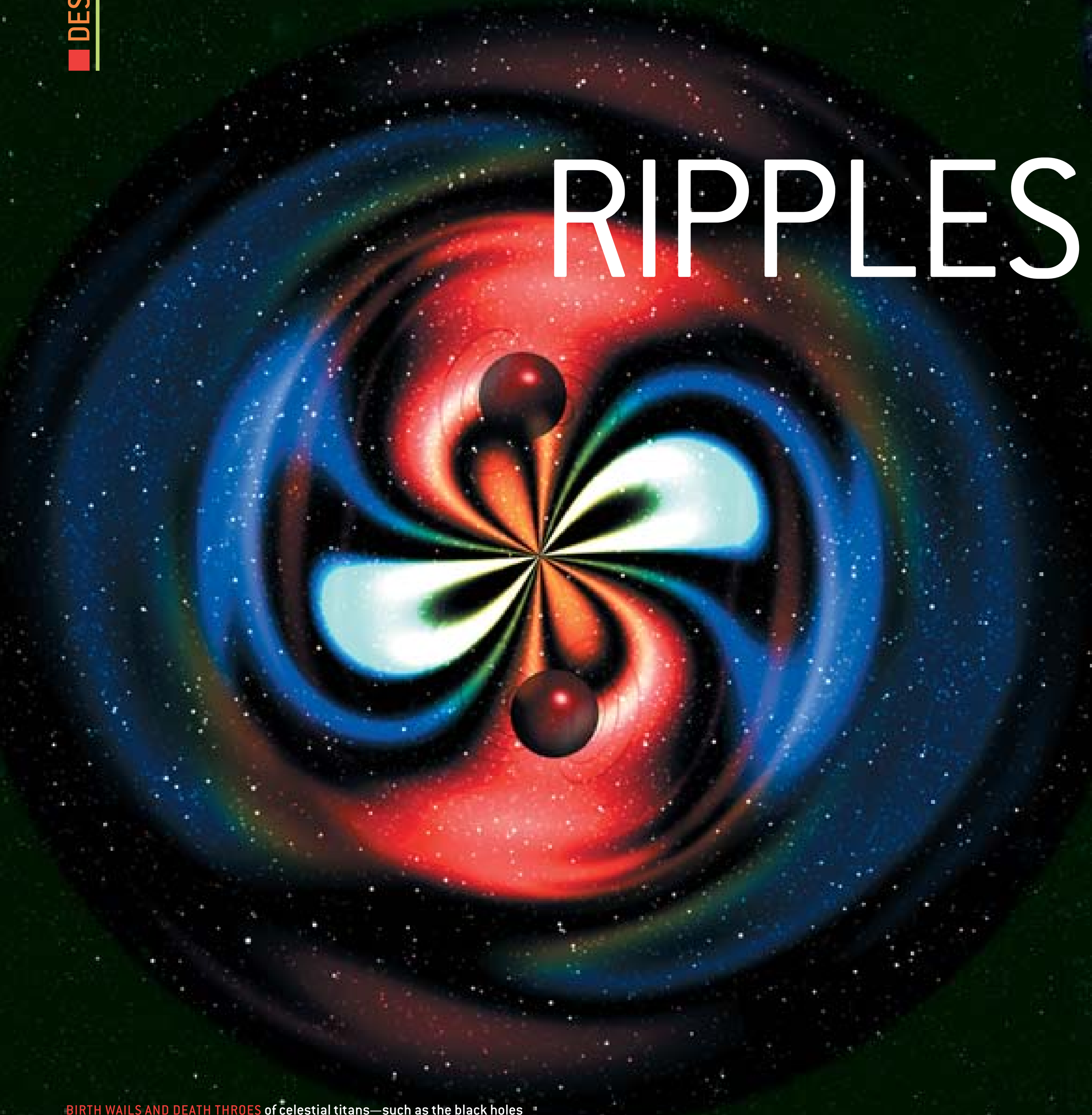
Cosmology is not only a fundamental science; it is also the grandest of the environmental sciences. How did a hot, amorphous fireball evolve, over 10 billion to 15 billion years, into our complex cosmos of galaxies, stars and planets? How did atoms assemble—here on Earth and perhaps on other worlds—into living beings intricate enough to ponder their own origins? These questions are a challenge for this millennium. Answering them may well be an unending quest. SA

MORE TO EXPLORE

Planet Quest: The Epic Discovery of Alien Solar Systems. Ken Croswell. Free Press, 1997.

The Little Book of the Big Bang: A Cosmic Primer. Craig J. Hogan. Copernicus, 1998.

RIPPLES



BIRTH WAILS AND DEATH THROES of celestial titans—such as the black holes (*spheres*) colliding in this supercomputer simulation—rumble through the universe on waves of gravitational energy. This year new instruments of astonishing size and sensitivity are trying to tune in those signals for the first time.

IN SPACETIME

PHYSICISTS HAVE SPENT EIGHT YEARS AND $365 MILLION BUILDING A RADICALLY NEW KIND OF OBSERVATORY TO DETECT GRAVITATIONAL WAVES. BUT WILL IT WORK? A TRIAL RUN PUT IT TO THE TEST

BY W. WAYT GIBBS

HANFORD, WASH., AND LIVINGSTON, LA.—A chill January wind sends a shiver through Frederick J. Raab as he stands, binoculars to his eyes, on a mound near the center of the LIGO Hanford Observatory. He runs his gaze northward down a ruler-straight concrete tunnel to a building four kilometers to the north: there is one end of the observatory. Pivoting 90 degrees, Raab pans westward across the sagebrush-stubbled desert until he spots an identical tube and another building, also four kilometers distant. "When we talk about locking the laser beam" that shines inside those tubes, Raab says, "we mean holding the light waves steady to better than the width of an atom—over that distance."

Raab oversaw the construction of this giant try square, one of a pair that are the largest, most expensive and—if they fulfill the ambition of their designers—most sensitive detectors yet to join the 40-year hunt for gravitational waves. Part ruler, part clock, these two instruments are spacetime meters that will attempt to record how the continuum is rattled by the most violent cataclysms in the universe: detonating stars, colliding black holes, perhaps phenomena not yet imagined. As these ripples expand outward at the speed of light, they alternately stretch and squeeze space, causing the distance between free-floating objects to expand and contract. But by the time the vibrations reach the earth, theorists estimate, they are so unsubstantial that they alter distances by less than one part in a trillion billion.

For all the cutting-edge technology crammed into LIGO, it is not yet clear whether it can attain that incredible sensitivity. Reduced to such a tiny murmur, the mightiest cosmic events are easily overpowered by the gentlest mundane disturbance. "The tide deforms the earth's crust as well as the oceans," Raab tells me. It moves the buildings here by a third of a millimeter, 100 billion times the displacement a gravitational wave would cause. Every earthquake in the world over magnitude six, the rumble of every truck on nearby roads, the computer fans in the lab next door—all these things shake the ground by more than an atom's width. "Even engine noise from jets passing overhead can work its way in," Raab says.

Down in the control room, we watch as the instrument struggles to compensate for the thumps and bumps. Fourteen days into an 18-day test run that began on December 28, the noise is winning. Raab stares at a panel of graphs projected onto the far wall. A red line bounces up and down, charting the status of the main detector here as it is thrown out of whack, steadies itself and gets knocked out again a few minutes later. A blue line that represents a smaller quality-control detector has gone flat altogether.

During a teleconference, physicist H. Richard Gustafson troubleshoots glitches with his counterparts at the LIGO Livingston Observatory, which sits in the backwoods of Louisiana. Joining the conversation is the director of the GEO 600, a similar but smaller instrument near Hannover, Germany. "Here at Hanford we had an awful night," Gustafson says, recounting troubles with computer crashes and noisy electronics.

The instrument in Louisiana has been more predictable. During the night it runs smoothly, but at 6:30 A.M. its line on the control screen goes flat as morning traffic picks up on Interstate 12 a few miles from the observatory and as Weyerhaeuser loggers begin felling loblolly pines nearby. GEO, with its shorter, 600-meter arms and less demanding precision, has been a model of reliability, on duty more than 90 percent of the time. But scientists need all three instruments up and running, and over two weeks the best stretch of simultaneous operation lasted just over an hour and a half.

Szabolcs Márka, a 32-year-old Hungarian postdoc at LIGO Livingston, seems content with the progress so far on this test run, the fifth that he has managed and the last before the two instruments were to begin routine round-the-clock observations in May. "As usual we are in problem-solving mode," Márka says.

Labor Pains, Death Gasps

EVER SINCE THE FOUNDERS of the LIGO project—Kip S. Thorne and Ronald Drever of the California Institute of Technology and Rainer Weiss of the Massachusetts Institute of Technology—first proposed the Laser Interferometer Gravitational Wave Observatory in 1984, no one has doubted that only a Herculean feat of engineering would make it work. That is one reason that "the project faced tremendous opposition from astronomers," says Harry M. Collins, a sociologist at Cardiff University in Wales who has studied the field's halting expansion from a backwater of physics to Big Science.

"The National Science Foundation turned down our first two proposals," Thorne remembers. "And the third, submitted in 1989, went through five years of very extensive review." High-profile astronomers, notably Jeremiah P. Ostriker of Princeton University, objected to the steep price, which by 1993 had risen to $250 million. They feared that smaller and less risky projects would get elbowed out of the budget. A blue-ribbon panel set up to rank U.S. astronomers' priorities for the 1990s excluded LIGO from its wish list. "It was a unanimous decision," recalls John Bahcall, an astrophysicist at the Institute for Advanced Studies in Princeton, N.J., who chaired the committee. Congress passed on the LIGO proposal at first, not approving funding until 1994.

Thorne and other proponents of LIGO argued that gravitational signals could launch a whole new field of astronomy, because they carry information about the universe that scientists can gather in no other way. These ethereal ripples were predicted in 1918 by Albert Einstein, who saw them as an unavoid-

DETECTING A QUIVER SO MINUSCULE IS LIKE NOTICING THAT SATURN HAS MOVED CLOSER TO THE SUN BY THE WIDTH OF A HYDROGEN ATOM.

Overview/*Gravitational-Wave Detectors*

- Although astronomers have never detected gravitational waves directly, Einstein's theory of relativity predicts that violent cataclysms such as black hole collisions will cause the fabric of space itself to vibrate.
- By the time they reach the earth, these ripples are so faint that picking them out of the surrounding noise is comparable to noticing a single grain of sand added to all the beaches of Long Island, N.Y.
- Six ultraprecise interferometers have been built around the world to detect these signals. Three are in the U.S. and began scientific observations in May. But they are still struggling to reach the necessary sensitivity.

PRECEDING PAGES: WERNER BENGER AEI/ZIB (*visualization*); NUMERICAL RELATIVITY GROUP AT AEI (*simulation/science*); LBNL/NERSC (*computing facility*); MAX PLANCK SOCIETY, EU ASTROPHYSICS NETWORK PROJECT, U.S. DEPARTMENT OF ENERGY (*support*)

CONTROL ROOM of the LIGO Livingston Observatory was home away from home for Caltech physicist Szabolcs Márka during the 18-day trial run, which he managed. Despite the challenges, the team was able to collect more than 70 hours of scientific data from all three U.S. interferometers at once.

MAX AGUILERA-HELLWEG

able consequence of his general theory of relativity. The attractive force we call gravity, Einstein famously postulated, occurs because massive bodies warp the four-dimensional fabric of the universe. If a dense object moves violently, space shudders in response.

When a giant star, for example, exhausts its fuel, it can detonate in a flash as luminous as 10 billion suns—a supernova. Astronomers believe that the star's outer layers are blown into space, while its iron core implodes with enough force to combine all its electrons and protons into neutrons and exotic particles. Within minutes, a solid metal sphere as big as the earth collapses into a neutron star less than 20 kilometers across. It is so dense that a teaspoonful of its surface would weigh nearly a billion tons. Scientists expect that a somewhat lopsided supernova would send out neutrinos and a burst of gravitational energy that would hit the earth several minutes before the flash arrived—time enough to alert conventional astronomers to train their telescopes on it. More important, details about the birth of the neutron star could be extracted from the gravitational signal even though the nascent object itself is tiny and swaddled in a blanket of fiery gas.

LIGO was designed to detect the death of neutron stars as well as their birth. Most stars orbit a mate, and occasionally both stars in a binary pair will go supernova yet remain locked in mutual thrall. With each revolution, the two neutron stars lose a little energy as they induce wrinkles in the surrounding fabric of space. Their orbit thus tightens step by step until they rip apart and merge, sometimes creating a black hole. Near the end of their frenetic tango, the massive bodies whirl around each other hundreds of times a second, flapping the bedsheets of spacetime around them. Radio pulses from such binary systems offer the most convincing, if indirect, evidence so far that gravitational waves actually exist.

But it is still anyone's guess whether the Caltech and M.I.T. groups that operate LIGO for the NSF will be able to detect such waves directly. "The curious thing about LIGO," Collins says, "is that, at least in its first instantiation, it cannot promise success."

Spectral Phenomena

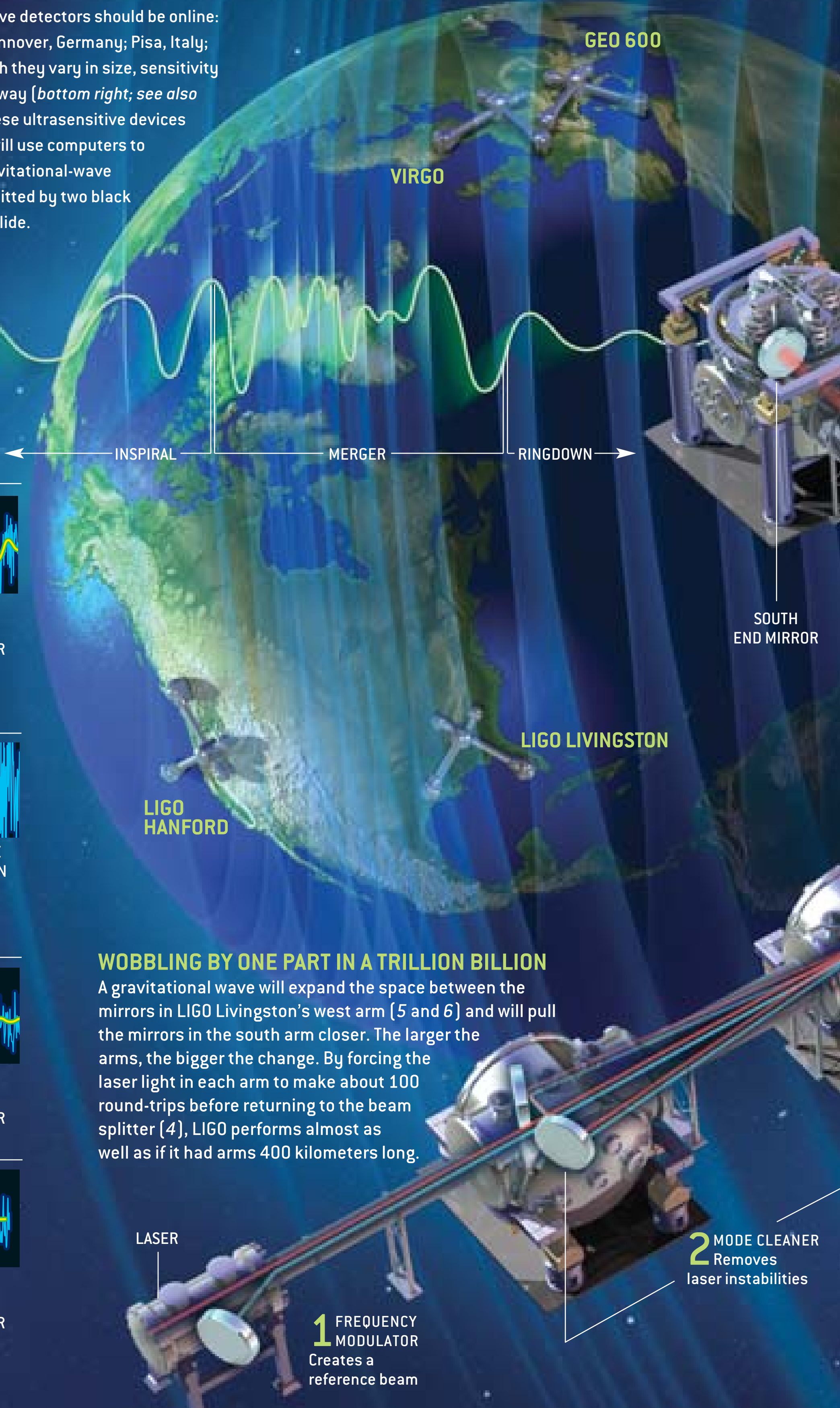
THE PROBLEM IS NOT that gravitational waves are weak. "The energy in gravitational waves is amazingly huge," says Gabriela I. González, a physicist at Louisiana State University. During the final minute that neutron stars spiral to their death 65 million light-years from the earth, the gravitational pulse would be so energetic that "if it arrived in the

GLOBAL GRAVITY OBSERVATORY

BY THE END OF 2003, six new gravitational-wave detectors should be online: one each near the cities of Livingston, La.; Hannover, Germany; Pisa, Italy; and Tokyo; and two at Hanford, Wash. Although they vary in size, sensitivity and details, all work in more or less the same way (*bottom right; see also box on page 94 for more details*). Because these ultrasensitive devices pick up so much terrestrial noise, scientists will use computers to scan the raw output (*below*) for predicted gravitational-wave patterns, such as the chirp, crash and ring emitted by two black holes in the seconds before and after they collide.

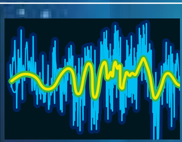
PINPOINTING THE SOURCE

As a gravitational pulse sweeps through the earth, the same waveform (*green*) will hit each detector at a slightly different time, allowing astronomers to pinpoint the source and eliminate other possible causes of the vibration.



LIGO

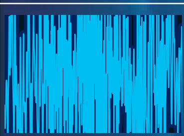
SPONSOR: U.S.
ARM LENGTH: 4 km at Livingston; 4 km and 2 km at Hanford
PEAK SENSITIVITY: Three parts in 10^{23} at 180 Hz
STATUS: Observations began in May 2002
COST: $530 million through 2007



MATCH WITH TEMPLATE, POSSIBLE MERGER

TAMA 300 (NOT SHOWN)

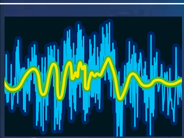
SPONSOR: Japan
ARM LENGTH: 300 m
PEAK SENSITIVITY: Five parts in 10^{21} from 700 to 1,000 Hz
STATUS: Preliminary observations began in 2001
COST: $10 million



OFFLINE BECAUSE OF EARTHQUAKE IN INDONESIA

GEO 600

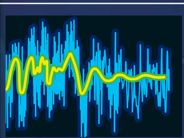
SPONSORS: U.K., Germany
ARM LENGTH: 600 m
PEAK SENSITIVITY: Eight parts in 10^{23} at 600 Hz
STATUS: Observations to begin in 2002
COST: $10 million



MATCH WITH TEMPLATE, POSSIBLE MERGER

VIRGO

SPONSORS: Italy, France
ARM LENGTH: 3 km
PEAK SENSITIVITY: One part in 10^{22} at 500 Hz
STATUS: Observations to begin in 2003
COST: $66 million



MATCH WITH TEMPLATE, POSSIBLE MERGER

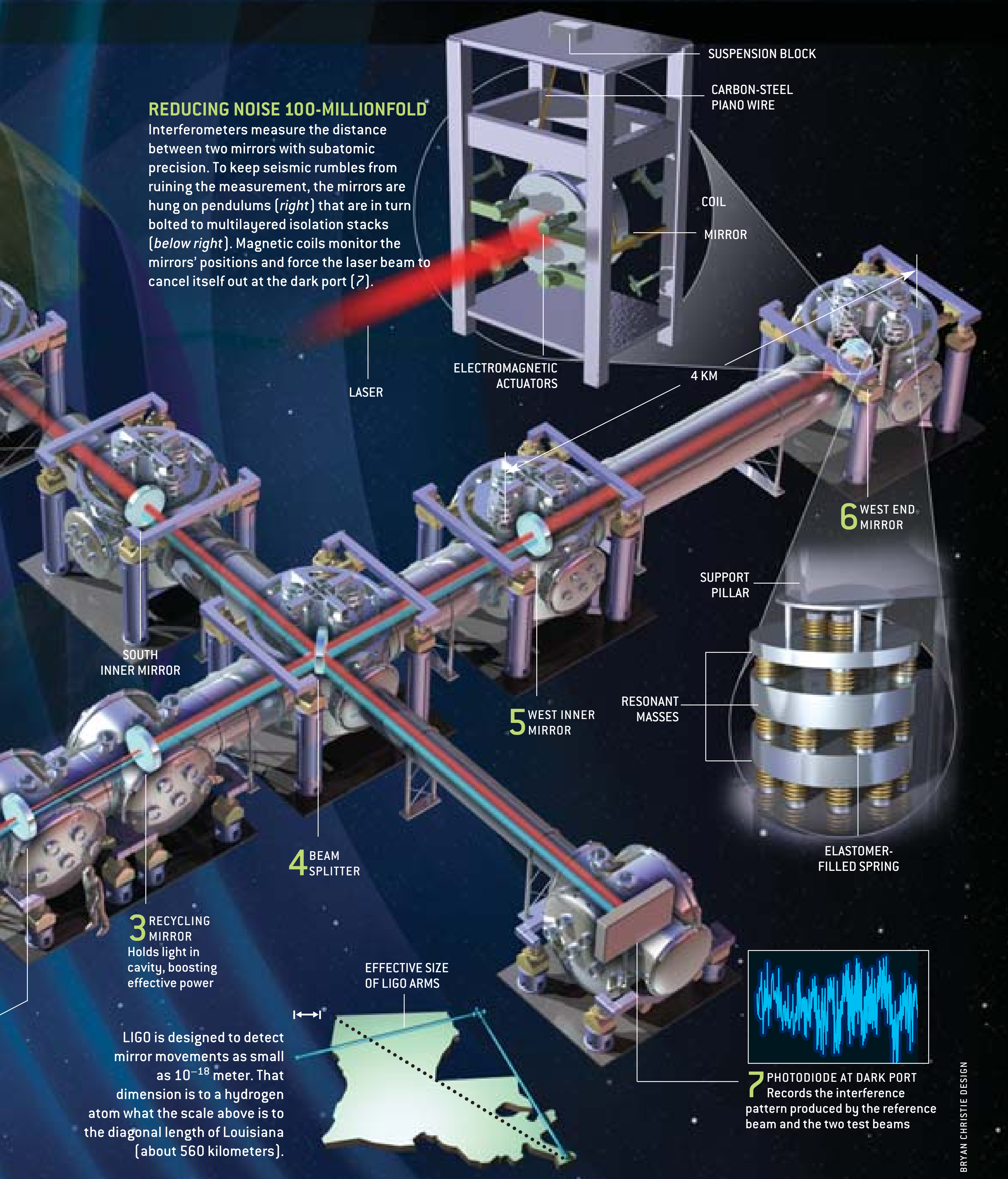
"Peak sensitivity" refers to design goals not yet achieved

WOBBLING BY ONE PART IN A TRILLION BILLION

A gravitational wave will expand the space between the mirrors in LIGO Livingston's west arm (*5* and *6*) and will pull the mirrors in the south arm closer. The larger the arms, the bigger the change. By forcing the laser light in each arm to make about 100 round-trips before returning to the beam splitter (*4*), LIGO performs almost as well as if it had arms 400 kilometers long.

REDUCING NOISE 100-MILLIONFOLD

Interferometers measure the distance between two mirrors with subatomic precision. To keep seismic rumbles from ruining the measurement, the mirrors are hung on pendulums (*right*) that are in turn bolted to multilayered isolation stacks (*below right*). Magnetic coils monitor the mirrors' positions and force the laser beam to cancel itself out at the dark port (7).



LIGO is designed to detect mirror movements as small as 10^{-18} meter. That dimension is to a hydrogen atom what the scale above is to the diagonal length of Louisiana (about 560 kilometers).

BRYAN CHRISTIE DESIGN

form of visible light, it would be brighter than the full moon," González says.

But unlike light, which deposits all its energy when it splats against matter, gravity passes ghostlike through solid objects with only a tingle of interaction. To a gravitational wave, the earth and everything on it are almost perfectly transparent. So even the powerful signal from the merging neutron stars will wiggle the center point of each mirror by just a few attometers (10^{-18} meter), the sensitivity that LIGO was designed to achieve.

As one arm of the observatory swells, the other will shrink. The phase and the frequency of the laser light inside the arms will shift in opposite directions. When the beams from the two arms are superimposed on a reference beam, they will be out of tune, and the wavering beats they generate can be decoded by computers to reveal the changing curvature of spacetime inside the arms. In principle, the technique, known as interferometry, can measure changes in distance much smaller than the wavelength of the infrared laser light—indeed, much smaller than the nucleus of an atom [*see box below*].

Ambitious as LIGO's sensitivity goal is, it leaves astronomers unimpressed. Neutron couplets are relatively rare; their deaths are spectacular but quick. Within 65 million light-years, astronomers estimate, only one such merger occurs every 10,000 years. "So although it is possible that we would see these waves," Thorne says, "it is not highly probable." He thinks it more likely that LIGO would pick up black hole mergers, which are 100 times more powerful than the neutron star variety. But theorists are uncertain by a factor of 1,000 how frequently these events

A Photon's Journey through LIGO

TO UNDERSTAND HOW THE LIGO interferometer works, imagine the adventure of a photon as it passes through the instrument. (We will neglect some details for clarity.) The photon is created in a suitcase-size laser that is as powerful as 20,000 laser pointers. It is one of trillions of photons marching in lockstep in an infrared beam.

1 Part of the beam takes a detour into a device that converts the light into two reference beams, one of slightly higher frequency than the main beam and one of slightly lower frequency. This frequency modulator thus creates a benchmark against which the test beam can be compared at the end of its journey. After the detour, the beams recombine and pass through a quartz window and into the first vacuum chamber. The builders took every precaution to prevent our photon from scattering out of its intended path. Vacuum pumps hold the air pressure below one trillionth of an atmosphere. The mirrors, as wide as dinner plates and 10 centimeters thick, have been polished to an accuracy of better than 16 atoms. And the thickness of reflective films coating the optics varies by no more than two atoms.

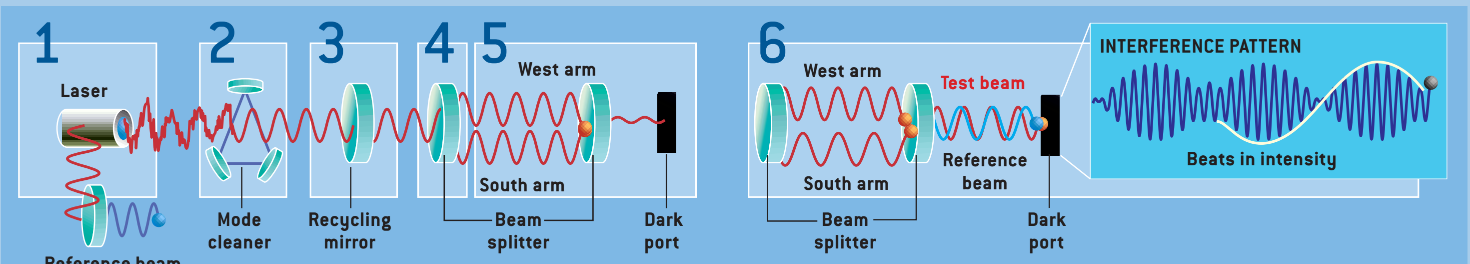
2 The photon enters a loop formed by three mirrors arranged in a narrow triangle. This mode cleaner is a quality checkpoint: the photon can move ahead only if its part of the beam has just the right shape and direction. Any light that is out of place or poorly aimed is tossed out a porthole.

3 The photon next zips through a half-silvered mirror, which blocks most photons that try to head back toward the laser. By trapping photons within the device, this mirror increases the power of the light beam 16-fold or more. Almost 100-fold additional amplification occurs in the instruments' long arms so that the beam inside those arms reaches the power of 20 million laser pointers.

4 At the beam splitter, the photon divides into identical twins. One stream of photons continues forward into the west arm. The other stream is inverted, its peaks flipped to valleys, as it is reflected into the south arm. The two test beams fly through the inner mirrors and into steel tubes four kilometers long. But the frequency-shifted reference beams are denied entry. They retrace their steps toward the beam splitter and circulate among the central optics until photons from the test beams return.

5 Meanwhile our photon and its inverted twin sail down the long arms to bounce off a mirror at each end. Although the atoms on the mirror surface are vibrating with heat, their motion is random and the beam hits trillions of atoms at once. On average the thermal vibrations nearly cancel out. The twin photons carom between the inner and end mirrors inside their respective arms. They make about 100 round-trips before leaking through the inner mirror and reuniting at the beam splitter, which sends them northward toward the dark port. Normally our photon and its alter ego will be at opposite points in their oscillation. Crest will meet trough, and the two will annihilate each other. The dark port will remain dark.

6 But if during the photons' journey, a gravitational wave slices through the apparatus, it will have curved space, lengthening one arm and shortening the other. Crest meets crest, and the dark port will light up. What's more, the reunited photons will combine with the frequency-modulated reference beams. Like musical notes played slightly out of tune, the light will beat, growing dimmer and brighter with the passage of the gravitational wave. Finally hitting a photodiode, the photon is converted into a perceptible electronic signal, the trace of a trembling spacetime. *—W.W.G.*



might occur within LIGO's range. There may be 10 a year or only one a century.

Going out to 300 million light-years would improve the odds, but then a typical event would change the relative length of LIGO's arms by only about one part in 10^{22}. Observers will have to wait for version 2.0 of LIGO to detect such a minuscule quiver, which is comparable to noticing that Saturn has moved closer to the sun by the width of a hydrogen atom.

The Unquiet Earth

AS IF THAT WERE not difficult enough, LIGO engineers must contend with the fact that mirrors wiggle for lots of reasons that have nothing to do with supernovae, neutron stars or black holes. Heat causes molecules in the mirrors and the wires on which they hang to jostle randomly. This thermal noise can drown out gravitational waves whose frequencies lie between 50 and 200 hertz. At higher frequencies, the interferometer is overwhelmed by the quantum effect called shot noise, which occurs because the number of photons hitting its sensors varies from one instant to the next. "You could turn the laser power up to boost the signal over the noise," explains Norna Robertson, one of the designers of the GEO instrument. "But if you put too much light in, it kicks the mirrors around in random ways."

At the moment, though, the biggest challenge for LIGO is at low frequencies, where the earth is constantly in motion. "At 100 hertz, the ground moves up and down by about 10^{-11} meter," Raab says. "We want to see motions of 10^{-19} meter" because that is a $10^{22\text{nd}}$ of the four-kilometer length of LIGO's arms. "So we need to reduce seismic noise 100-millionfold."

We put on goggles and shoe covers and head over to the high bay that contains the laser and most of the detector's sensors. As he opens the door to the cavernous room, Raab lowers his voice to just above a whisper. I try to tread gently.

Raab walks over to a steel vacuum chamber as big as an upended van. To get from the ground to the mirror inside, a seismic rumble must pass through a stack of devices designed to sap its energy: a one-meter slab of reinforced concrete, scissor jacks, air bearings, four layers of thick custom-made springs, four heavy steel plates (each resonating at a different frequency) and, finally, a pendulum of steel piano wire. "We reduce seismic noise by a factor of 100 in the pendulum suspension and by another factor of a million with the isolation stacks," Raab notes. Some ground movements, such as lunar tides, still must be fought with more active devices, such as computer-controlled electromagnets that push and pull on tiny magnets glued to the mirrors.

Yet sometimes dampening external noise 100-millionfold isn't enough. "Just recently there was a magnitude-seven earthquake in Sumatra; that knocked us offline," Raab says. Strong winds have pulled the Hanford interferometer out of lock as well.

Not all seismic noise is natural. Robert Schofield, a postdoc at the University of Oregon, has become the noise detective for LIGO. One evening he is sitting at a control station frowning at a chart of the latest signals picked up by the detector. "Look at this peak," he says. "Right here at 2.3 hertz. I hadn't noticed it before because it is so narrow, but it accounts for 20 percent of the noise getting into the interferometer." Scanning over readouts from a battery of seismometers that surround the observatory, he concludes that the noise is coming from near the 200 East section of the Hanford Nuclear

MARK COLES, director of LIGO's Louisiana facility, is trying to deal with logging, traffic and other sources of noise that thwart his engineers' efforts. "It may be that in the first few years we won't be able to get all the way to full design sensitivity," he says. "But it's still a great project to work on."

BRYAN CHRISTIE DESIGN (*opposite page*); MAX AGUILERA-HELLWEG (*above*)

Reservation, the 1,400-square-kilometer radioactive waste depository that surrounds the LIGO Hanford site.

Schofield marches down the hall, grabs a seismometer and an oscilloscope and hauls them into a van. He drives several miles farther into the reservation, then pulls over and sets up his equipment. We can see the bright lights of some night operation in 200 East several miles away. But we can't get any closer because the area contains tanks of plutonium-laced waste, and it is protected by security forces with submachine guns. Schofield sets the seismometer to listen for almost five minutes. But it reveals no trace of the 2.3-hertz noise. "I think it must be a large piece of rotating machinery doing some fiendish thing out there," he tells co-workers later.

Fortunately, noise near two hertz is not an immediate problem. LIGO, like the other giant gravitational-wave observatories nearing completion—GEO in Germany, TAMA in Tokyo and VIRGO near Pisa, Italy—is tuned to listen for gravitational waves from only 40 to about 3,000 hertz, coincidentally right in the range of human hearing. In the control room, LIGO operators have connected a speaker to sensors on the interferometer; it plays what the device "hears." A nearby supernova might come through as a burst of static. The wail of dying neutron stars would start low and sweep higher in an almost musical chirp.

Noise usually hisses and pops, but occasionally some recognizable sound leaks in. "There is a periscope on the laser table that raises the beam up to the right height," Schofield explains. Various noises can shake the periscope, introducing subtle Doppler shifts in the frequency of light passing through it. "If someone is talking near that periscope," he says, "you can hear their voice on that speaker."

THEORISTS HAVE A VERY POOR TRACK RECORD FOR PREDICTING WHAT WE WILL SEE WHEN A NEW WINDOW IS OPENED ON THE UNIVERSE.

So in addition to the seismographs, LIGO engineers have studded the facility with microphones and magnetometers, as well as sensors that monitor temperature, pressure and wind. A stream of data from about 5,000 sensor channels gets recorded simultaneously. The first thing that scientists would do if they thought they saw a gravitational wave is look for glitches or noise that had leaked into the system.

On the last day of the test run, González hands the director, Mark Coles, a plot of the interferometer output from that morning. It contains a bounce that

Next-Generation Detectors

IF LIGO ACHIEVES the sensitivity for which it was designed, it will still have only a middling chance of detecting gravitational waves. "But our strategy from the beginning has been to do this in two steps," says Caltech physicist Kip S. Thorne: first get the machines working and gain confidence in their reliability, then upgrade to advanced components that will virtually guarantee regular signal detections.

Although the project's leaders have not yet made a formal proposal, they know roughly what they want. "It'll cost on the order of $100 million, begin around 2006 and take about two years to complete," says LIGO director Barry Barish. The laser will be boosted from 10 to 180 watts. Instead of single loops of steel wire, the optics will hang on silica ribbons attached to a three-stage pendulum now being tested in the GEO 600 detector in Germany. And the 11-kilogram silica glass mirrors will be replaced with 30-kilogram sapphire crystals.

The changes will boost sensitivity by a factor of 20, Barish estimates. That will put the instrument, Thorne says, "into the domain where, for the first time in history, humans will be seeing human-size objects behaving quantum-mechanically." Researchers have devised so-called quantum nondemolition techniques that can make measurements twice as precise as normally allowed by the Heisenberg indeterminacy principle. If it all works, "it will increase by 8,000 times the volume of space we can search," Barish says.

The Japanese have also designed a successor to their TAMA 300-meter interferometer, although project manager Yoshihide Kozai says, "I am afraid it will be a few years before we obtain the funds" to start construction. The Large-Scale Cryogenic Gravitational Wave Telescope would have three-kilometer arms built deep underground in the Kamioka mine. Supercooled sapphire mirrors of 51 kilograms each would help it match the sensitivity of LIGO II at frequencies below 40 hertz.

NASA and the European Space Agency are designing an even more ambitious gravitational-wave observatory called LISA. A trio of laser-toting satellites to be launched in 2011 would form an interferometer with arms five million kilometers long—better than 10 times the distance from the earth to the moon. As they orbited the sun, the trio would hold their position relative to one another with one-micron precision. LISA would be no more sensitive than LIGO II, but it could sense gravitational waves of much lower frequency than any detector built on the quaking earth.

"The most likely thing LISA would see is the motion of extremely massive black holes—from a million to billions of times the mass of the sun—orbiting each other in the center of very distant galaxies," Thorne says. "The astronomers are all over themselves about LISA," reports M.I.T.'s Rainer Weiss. "They know for sure they will see events." But because its cost will probably exceed half a billion dollars, Weiss predicts that "it will be much tougher to get LISA through Congress even than it was to get LIGO approved." *—W.W.G.*

looks like a real signal. It isn't. "We just invented a speedometer for the cattle guard on the entrance road," she says with a laugh. As each axle of a passing truck hits the horizontal rails, a rumble appears in the gravitational-wave channel.

Spurious signals can also be rejected by comparing data from two or more observatories, Márka explains. "If both LIGO sites see the same shape signal within a few milliseconds of each other and so does GEO, which sits on a different continental plate and is connected to a different electrical grid, then it is very, very unlikely to be a fake signal from some common source of noise."

Yet there is only so much they can do to overcome human-generated noise. The problem is especially bad at the Livingston site. "We can see the trains that go by three times a day," Coles says. "We can see the workers hauling trees. We can see when traffic picks up at lunchtime." During this test run, the Livingston instrument was online just 62 percent of the time, not including short blips. All three LIGO interferometers were up simultaneously for only 18 percent of the run.

"We know we have a problem with ground noise at Livingston," acknowledges Rainer Weiss, spokesperson for the LIGO Scientific Collaboration. "And it will get worse still. Society creeps in on us." Barry Barish, head of the project, says new active isolation stacks are being developed and will be installed next year. "I wish we didn't have to do it," Weiss says. "That was an engineering enhancement we had planned to add during the upgrade to LIGO II in 2006." The upgrade will add at least $750,000 to the $365 million that NSF has spent so far on the project and to the $165 million that it has just allotted for the next five years.

Yet even when the systems are locked and working, Weiss says, "we're still miles away—a factor of 1,000 away—from our design sensitivity. We're hoping to get at least 10 times better by June. But beyond that I don't know. There are many things we can try."

The uncertainty still troubles LIGO's old critic, Ostriker: "I have always believed that detecting gravitational waves will provide us insights obtainable in no other way. That said, I think that the LIGO program has been an egregious waste of funds—funds that could have been used for more productive science."

But Thorne sees things differently. "Theorists have a very poor track record for predicting what we will see when a new window is opened on the universe," he says. "Early radio telescopes discovered that the signals were much stronger than theorists expected. That happened again when the x-ray window opened in the 1960s. And when we started looking for neutrinos arriving from the sun, we were surprised by how few there were. In some sense, opening the gravitational window will give us a more radically different view on the universe than those previous advances did." Ripples in spacetime may shake up science yet. SA

W. Wayt Gibbs is senior writer.

STRAIGHTER THAN THE SURFACE OF THE EARTH, the concrete tunnel that houses the west arm of the LIGO Livingston Observatory rises more than a meter off the ground on its four-kilometer run as the planet curves beneath it. The tunnel houses an airtight steel pipe. Inside the pipe is a vacuum, and through the vacuum shines a beam of infrared light with the power of 20 million laser pointers.

MAX AGUILERA-HELLWEG

MORE TO EXPLORE

LIGO's main Web site is **www.ligo.caltech.edu**

Einstein's Unfinished Symphony. Marcia Bartusiak. Joseph Henry Press, 2000.

Laser Interferometric Gravitational Wave Detectors. Norna A. Robertson in *Classical and Quantum Gravity,* Vol. 17, No. 15, pages R19–R40; August 7, 2000. Available at **www.iop.org/Journals/CQG**

New Physics and Astronomy with the New Gravitational-Wave Observatories. Scott A. Hughes et al. in *Proceedings of the 2001 Snowmass Meeting.* Available at **www.ligo.caltech.edu/docs/P/P010029-00.pdf**

DESTINY

If the new cosmology fails, what's the backup plan?

plan B for the Cosmos

BY JOÃO MAGUEIJO

Although cosmic inflation has acquired an aura of invincibility, alternative theories explaining the evolution of the universe continue to attract some interest among cosmologists. The steady state theory, which until the 1960s was widely regarded as the main alternative to the big bang, has been kept alive by a small band of proponents. The pre–big bang theory, a reworking of inflation that has been motivated by string theory, also turns some heads. But the most promising and provocative alternative may be the varying-speed-of-light theory (VSL), which my colleagues and I have been developing for several years. If nothing else, these dissenting views add color and variety to cosmology. They also give expression to a nagging doubt: Could the enthusiasm generated by inflation and its offshoots conceal a monstrous error?

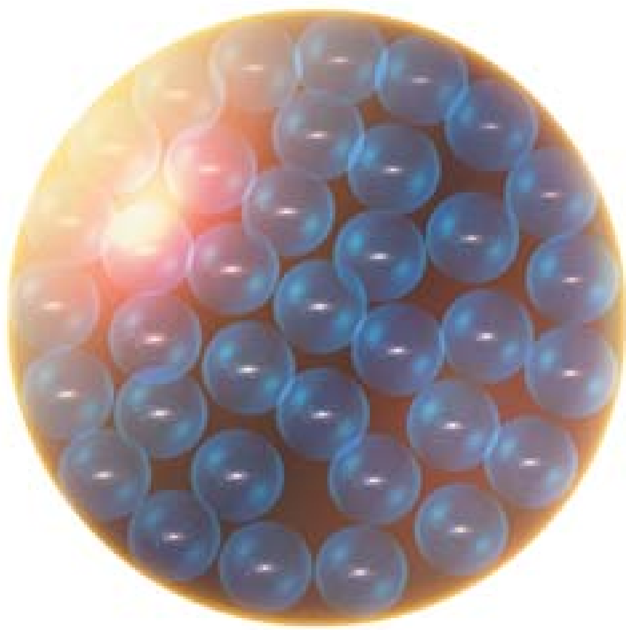
Mainstream cosmological theories such as inflation are based on a crucial assumption: that the speed of light and other fundamental physical parameters have had the same values for all time. (They are, after all, known as constants.) This assumption has forced cosmologists to adopt inflation and all its fantastic implications. And sure enough, experiments show that the presumed constants are not aging dramatically. Yet researchers have probed their values only over the past billion years or so. Postulating their constancy over the entire life of the universe involves a massive extrapolation. Could the presumed constants actually change over time in a big bang universe, as do its temperature and density?

Theorists find that some constants are more agreeable than others to giving up their status. For instance, the gravitational constant, G, and the electron's charge, e, have often been subjected to this theoretical ordeal, causing little scandal or uproar. Indeed, from Paul Dirac's groundbreaking work on varying constants in the 1930s to the latest string theories, dethroning the constancy of G has been exquisitely fashionable. In contrast, the speed of light, c, has remained inviolate. The reason is clear: the constancy of c and its status as a universal speed limit are the foundations of the theory of relativity. And relativity's spell is so strong that the constancy of c is now woven into all the mathematical tools available to the physicist. "Varying c" is not even a swear word; it is simply not present in the vocabulary of physics.

Yet it might behoove cosmologists to expand their vernacular. At the heart of inflation is the so-called horizon problem of big bang cosmology, which stems from a simple fact: at any given time, light—and hence any interaction—can have traveled only a finite distance since the big bang. When the universe was one year old, for example, light could have traveled just one light-

THE AUTHOR

JOÃO MAGUEIJO is a lecturer in theoretical physics at Imperial College, London. His research interests have oscillated between the observational and the "lunatic" (his word) aspects of cosmology, from analysis of the cosmic microwave background radiation to topological defects, quintessence and black holes. He is the author of the forthcoming book *Faster Than the Speed of Light* (Perseus Publishing).



TROUBLE ON THE HORIZON

At the stripling age of one year, the universe was subdivided into isolated pockets, demarcated by "horizons" one light-year in radius (*blue spheres*). Today the horizon is about 15 billion light-years in radius (*red sphere*), so it takes in zillions of these pockets. The odd thing is that despite their initial isolation, all the pockets look pretty much the same. Explaining this mysterious uniformity is the great success of the theory of inflation.

ILLUSTRATIONS BY ALFRED T. KAMAJIAN

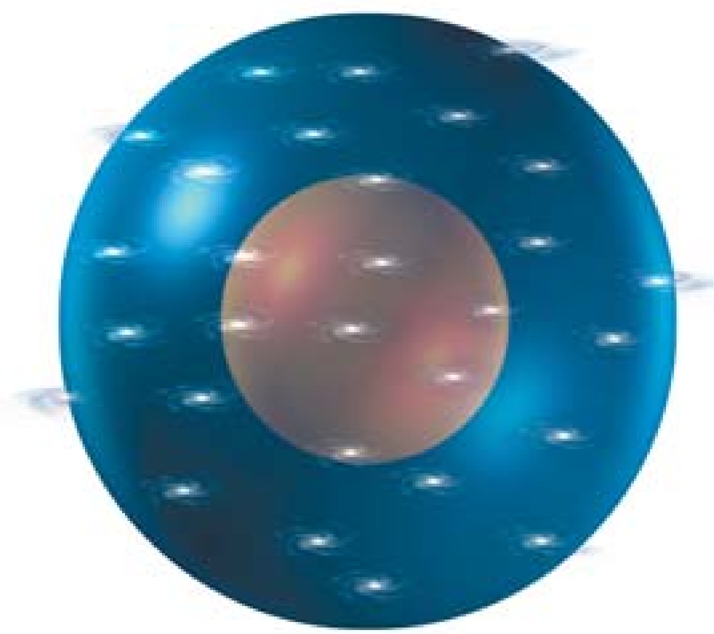
Updated from the January 2001 issue

year (roughly). The universe is therefore fragmented into horizons, which demarcate regions that cannot yet see one another.

The shortsightedness of the universe is enormously irritating to cosmologists. It precludes explanations based on physical interactions for puzzles such as why the early universe was so uniform. Within the framework of the standard big bang theory, the uniformity can be explained only by fine-tuning the initial conditions—essentially a recourse to metaphysics.

Inflation cunningly gets around this problem. Its key insight is that for a light wave in an expanding universe, the distance from the starting point is greater than the distance traveled. The reason is that expansion keeps stretching the space already covered. By analogy, consider a driver who travels at 60 kilometers an hour for one hour. The driver has covered 60 kilometers, but if the road itself has elongated in the meantime, the distance from the point of departure is greater than 60 kilometers. Inflationary theory postulates that the early universe expanded so fast that the range of light was phenomenally large. Seemingly disjointed regions could thus have communicated with one another and reached a common temperature and density. When the inflationary expansion ended, these regions began to fall out of touch.

It does not take much thought to realize that the same thing could have been achieved if light simply had traveled faster in the early universe than it does today. Fast light could have stitched together a patchwork of otherwise disconnected regions. These regions could then have homogenized themselves. As the speed of light slowed, those regions would have fallen out of contact.



BROADENING THE HORIZON

Inflation is not the only answer to the horizon problem. Instead maybe conditions in the early universe allowed light to travel faster than its present speed—a billion times faster or more. Zippy light made for bigger pockets (*blue sphere*). As light slowed to its present speed, the horizon shrank (*red sphere*). As a result, we now see just a part of one of the initial pockets, so it is no longer a mystery why the universe looks so uniform.

This was the initial insight that led Andreas C. Albrecht, then at the University of California at Berkeley, John Barrow of the University of Cambridge and me to propose the VSL theory. Contrary to popular belief, our motivation was not to annoy the proponents of inflation. (Indeed, Albrecht is one of the fathers of inflationary theory.) We felt that the successes and shortcomings of inflation would become clearer if an alternative existed, no matter how crude.

Naturally, VSL requires rethinking the foundations and language of physics, and for this reason a number of implementations are possible. What we first proposed was a reckless act of extreme violence against relativity, albeit with the redeeming merit of solving many puzzles besides the flatness problem. For example, our theory accounts for the minuscule yet nonzero value of the cosmological constant in today's universe. The reason is that the vacuum-energy density represented by the cosmological constant depends very strongly on c. A suitable drop in c reduces the otherwise domineering vacuum energy to innocuous levels. In standard theories, on the other hand, the vacuum energy cannot be diluted.

But our formulation is just one possibility, and the urge to reconcile VSL to relativity is motivating much ongoing work. The more cautious implementations of VSL pioneered by John Moffat of the University of Toronto and later by Ian T. Drummond of Cambridge are easier for relativity theorists to swallow. It now appears that the constancy of c is not so essential to relativity after all; the theory can be based on other postulates. Some have pointed out that if the universe is a three-dimensional membrane in a higher-dimensional space, as string theory suggests, the apparent speed of light in our world could vary while the truly fundamental c remains constant. It has also been suggested that a varying speed of light may be part and parcel of any consistent theory of quantum gravity.

Whether nature chose to inflate or to monkey with c can only be decided by experiment. The VSL theory is currently far less developed than inflation, so it has yet to make firm predictions for the cosmic microwave background radiation. On the other hand, some experiments have indicated that the so-called fine structure constant may not be constant. Varying c would explain those findings.

It remains to be seen whether these observations will withstand further scrutiny; meanwhile VSL remains a major theoretical challenge. It distinguishes itself from inflation by plunging deeper into the roots of physics. For now, VSL is far from being mainstream. It is a foray into the wild. SA

MORE TO EXPLORE

Time Varying Speed of Light as a Solution to Cosmological Puzzles. Andreas Albrecht and João Magueijo in *Physical Review D,* Vol. 59, No. 4, Paper No. 043516; February 15, 1999. Preprint available at **xxx.lanl.gov/abs/astro-ph/9811018**

Big Bang Riddles and Their Revelations. João Magueijo and Kim Baskerville in *Philosophical Transactions of the Royal Society A,* Vol. 357, No. 1763, pages 3221–3236; December 15, 1999. **xxx.lanl.gov/abs/astro-ph/9905393**

Lorentz Invariance with an Invariant Energy Scale. João Magueijo and Lee Smolin in *Physical Review Letters,* Vol. 88, No. 19; April 26, 2002.