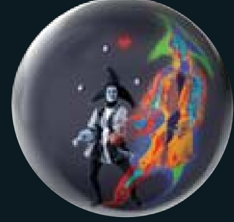
Colossal Galactic Explosions
Where matter turns to energy p. 50

Quarks, Gluons and Strings
Physics at the boundary p. 74

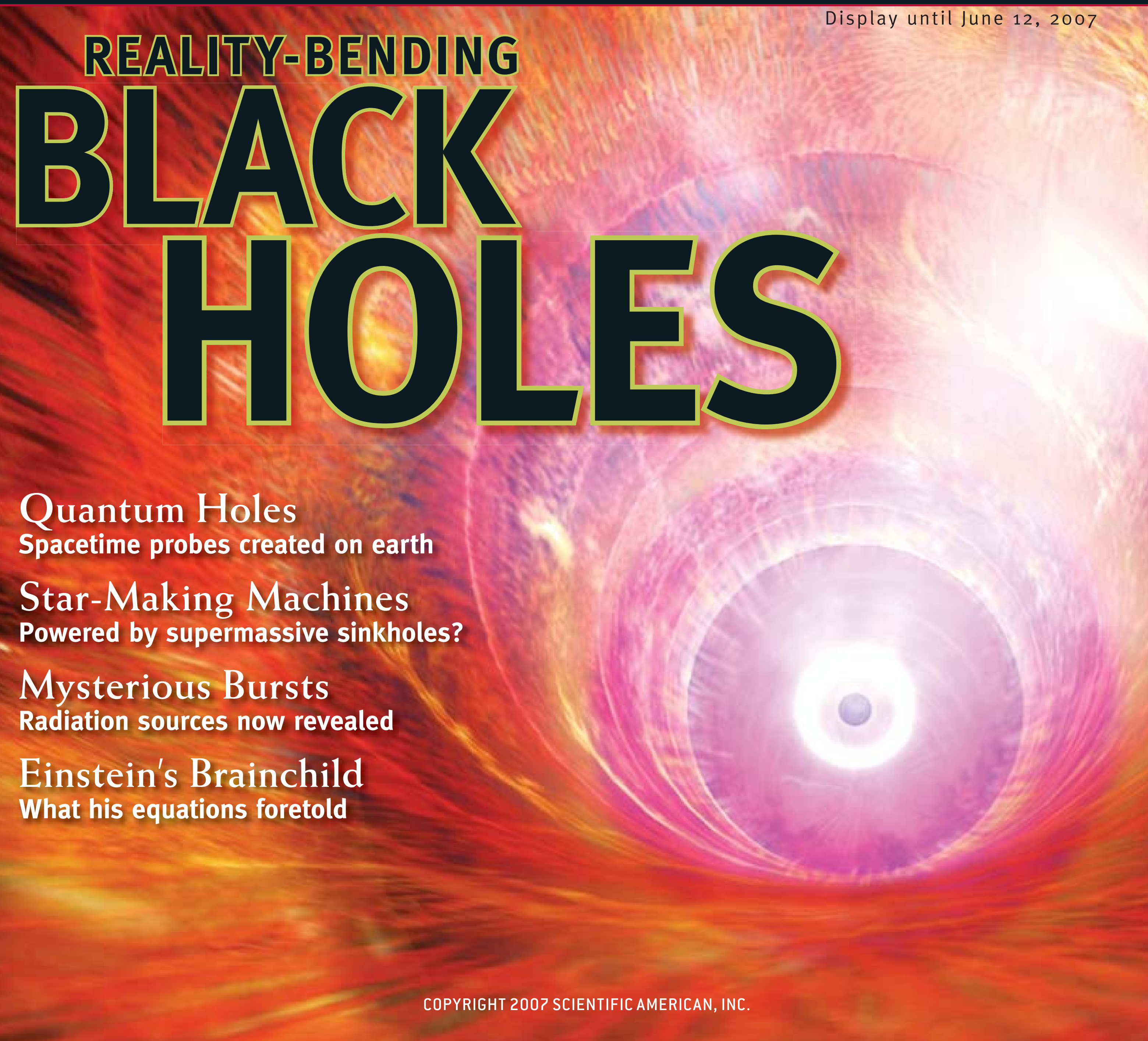


The Ultimate Time Warp
A portal to the future p. 28

SCIENTIFIC AMERICAN REPORTS

SPECIAL EDITION ON ASTROPHYSICS

www.sciam.com
Display until June 12, 2007



REALITY-BENDING BLACK HOLES

Quantum Holes
Spacetime probes created on earth

Star-Making Machines
Powered by supermassive sinkholes?

Mysterious Bursts
Radiation sources now revealed

Einstein's Brainchild
What his equations foretold

Page Intentionally Blank

SCIENTIFIC AMERICAN
REPORTS

Scientific American Reports is published by the staff of SCIENTIFIC AMERICAN, with project management by:

letter from the editor

Drawn to the Abyss

DON DIXON

BLACK HOLES curve the fabric of spacetime so extremely that it rends. The superdense objects devour anything—even light—that strays too close, a trip from which there is no escape. Perhaps their most singular power, however, is their hold on our imagination. Learning more about these implacable gluttons offers the same shivery frisson as watching a stalking horror-movie creature while knowing we are safe in our cushioned seats.

As the authors in this special issue explain, black holes offer much more to science than the can't-look-can't-look-away spectacle of destruction. The forces they unleash shape the regions around them, providing clues to the evolution of stars and galaxies. For instance, the dark sinkholes reveal a surprising bright side. In their quest to solve an enduring mystery, astronomers have learned that black holes are responsible for some of the most dazzling fireworks in the universe. When a massive star collapses to birth a black hole, it releases a titanic pulse of radiation in a gamma-ray burst that can be seen from billions of light-years away, as Neil Gehrels, Luigi Piro and Peter J. T. Leonard discuss in their article, "The Brightest Explosions in the Universe," starting on page 34. Greedily feeding supermassive black holes also exist in regions called starbursts, where stars are forming at a phenomenal rate. How? Turn to page 42 for "The Galactic Odd Couple," by Kimberly Weaver.

Studying black holes yields insights into other mind-bending areas of physics. In the coming years the highest-energy particle accelerators on earth might be able to produce distant cousins of the astrophysical behemoths: microscopic black holes. They would explode immediately after they formed, giving clues about how spacetime is woven together and whether it has unseen higher dimensions, explain Bernard J. Carr and Steven B. Giddings in "Quantum Black Holes." The article starts on page 20. Still other features in the issue explore what black holes can tell us about time travel, the nature of gravity, the ultimate amount of information the universe can hold and whether our seemingly 3-D reality is actually an illusion. So draw up that comfortable chair and get ready to learn more about one of the most awesome beasts in the universe.

Mariette DiChristina
Executive Editor
Scientific American
editors@sciam.com

12

28

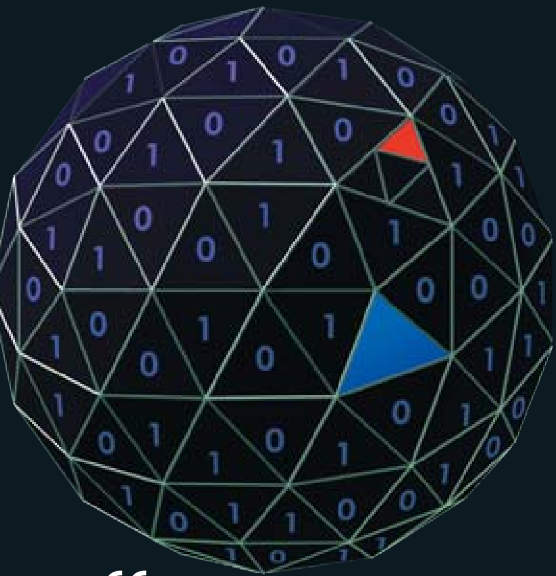
50

Contents

Volume 17, Number 1, 2007
SCIENTIFIC AMERICAN
REPORTS

Cover illustration by DON DIXON

The articles in this special edition are updated from previous issues of *Scientific American*.



66

74

82

Scientific American Special (ISSN 1048-0943), Volume 17, Number 1, 2007, published by Scientific American, Inc., 415 Madison Avenue, New York, NY 10017-1111.

Canadian BN No. 127387652RT; QST No. Q1015332537. To purchase additional quantities: U.S., $10.95 each; elsewhere, $13.95 each. Send payment to Scientific American, Dept. BH2007, 415 Madison Avenue, New York, NY 10017-1111. **Inquiries: fax 212-355-0408 or telephone 212-451-8442.** Printed in U.S.A.

the reluctant FATHER OF BLACK HOLES

Albert Einstein's equations of gravity are the foundation of the modern view of black holes; ironically, he used the equations in trying to prove these objects cannot exist

By Jeremy Bernstein

PRO AND CON: In 1939 J. Robert Oppenheimer (*right*) argued for the existence of black holes, at the same time Albert Einstein tried to disprove them. Their careers crossed paths at the Institute for Advanced Study in Princeton, N.J., in the late 1940s, when this photograph was taken, but it is unknown whether they ever discussed black holes.

CORBIS

Great science sometimes produces a legacy that outstrips not only the imagination of its practitioners but also their intentions. A case in point is the early development of the theory of black holes and, above all, the role played in it by Albert Einstein. In 1939 Einstein published a paper in the journal *Annals of Mathematics* with the daunting title "On a Stationary System with Spherical Symmetry Consisting of Many Gravitating Masses." With it, Einstein sought to prove that black holes—celestial objects so dense that their gravity prevents even light from escaping—were impossible.

The irony is that, to make his case, he used his own general theory of relativity and gravitation, published in 1916—the very theory that is now used to argue that black holes are not only possible but, for many astronomical objects, inevitable. Indeed, a few months after Einstein's rejection of black holes appeared—and with no reference to it—J. Robert Oppenheimer and his student Hartland S. Snyder published a paper entitled "On Continued Gravitational Contraction." That work used Einstein's general theory of relativity to show, for the first time in the context of modern physics, how black holes could form.

Einstein sought to prove that **black holes were impossible.**

Perhaps even more ironically, the modern study of black holes, and more generally that of collapsing stars, builds on a completely different aspect of Einstein's legacy—namely, his invention of quantum-statistical mechanics. Without the effects predicted by quantum statistics, every astronomical object would eventually collapse into a black hole, yielding a universe that would bear no resemblance to the one we actually live in.

Overview
Black Hole History

- Using his general theory of relativity, Albert Einstein sought to prove that "Schwarzschild singularities" (which later became known as black holes) cannot exist. In a 1939 paper he tried to kill the idea once and for all.
- Around the same time, J. Robert Oppenheimer and his student Hartland S. Snyder used Einstein's general theory of relativity to show that a collapsing star of sufficient mass could form a black hole.
- The modern study of black holes builds not only on the general theory of relativity but also on other work done by Einstein.

Bose, Einstein and Statistics

EINSTEIN'S CREATION of quantum statistics was inspired by a letter he received in June 1924 from a then unknown young Indian physicist named Satyendra Nath Bose. Along with Bose's letter came a manuscript that had already been rejected by one British scientific publication. After reading the manuscript, Einstein translated it himself into German and arranged to have it published in the prestigious journal *Zeitschrift für Physik*.

Why did Einstein think that this manuscript was so important? For two decades he had been struggling with the nature of electromagnetic radiation—especially the radiation trapped inside a heated container that attains the same temperature as its walls. At the start of the 20th century German physicist Max Planck had discovered the mathematical function that describes how the various wavelengths, or colors, of this "black body" radiation vary in intensity. It turns out that the form of this spectrum does not depend on the material of the container walls. Only the temperature of the radiation matters. (A striking example of black-body radiation is the photons left over from the big bang, in which case the entire universe is the "container." The temperature of these photons has been measured at 2.726 ± 0.002 kelvins.)

Somewhat serendipitously, Bose had worked out the statistical mechanics of black-body radiation—that is, he derived the Planck law from a mathematical, quantum-mechanical perspective. That outcome caught Einstein's attention. But being Einstein, he took the matter a step further. He used the same methods to examine the statistical mechanics of a gas of massive molecules obeying the same kinds of rules that Bose had used for the photons. He derived the analogue of the Planck law for this case and noticed something absolutely remarkable. If one cools the gas of particles obeying so-called Bose-Einstein statistics, then at a certain critical temperature all the molecules suddenly collect themselves into a "degenerate," or single, state. That state is now known as Bose-Einstein condensation (although Bose had nothing to do with it).

An interesting example is a gas made up of the common isotope helium 4, whose nucleus consists of two protons and two neutrons. At a temperature of 2.18 kelvins, this gas turns into a liquid that has the most uncanny properties one can imagine, including frictionless flow (that is, superfluidity). More than a decade ago U.S. researchers accomplished the difficult task of cooling other kinds of atoms to several billionths of a kelvin to achieve a Bose-Einstein condensate.

Not all the particles in nature, however, show this condensation. In 1925, just after Einstein published his papers on the condensation, Austrian-born physicist Wolfgang Pauli identified a second class of particles, which includes the electron, proton and neutron, that obey different properties. He found that no two such identical particles—two electrons, for example—can ever be in exactly the same quantum-mechanical state, a property that has since become known as the Pauli exclusion principle. In 1926 Enrico Fermi and P.A.M. Dirac invented the quantum statistics of these particles, making them the analogue of the Bose-Einstein statistics.

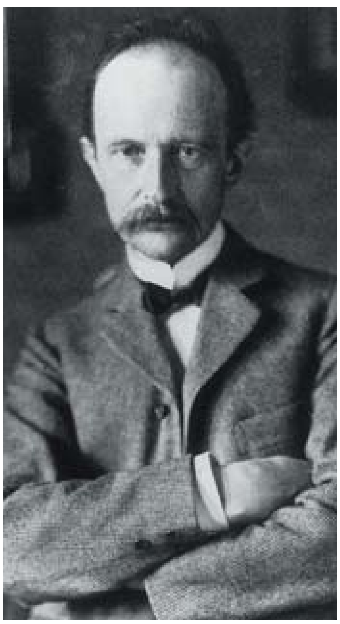
Because of the Pauli principle, the last thing in the world these particles want to do at low temperatures is to condense. In fact, they exhibit just the opposite tendency. If you compress, say, a gas of electrons, cooling it to very low temperatures and shrinking its volume, the electrons are forced to begin invading one another's space. But Pauli's principle forbids this, so they dart away from one another at speeds that can approach that of light. For electrons and the other Pauli particles, the pressure created by these fleeing particles—the "degeneracy pressure"—persists even if the gas is cooled to absolute zero. It has nothing to do with the fact that the electrons repel one another electrically. Neutrons, which have no charge, do the same thing. It is pure quantum physics.

Quantum Statistics and White Dwarfs

BUT WHAT HAS quantum statistics got to do with the stars? Before the turn of the century, astronomers had begun to identify a class of peculiar stars that are small and dim: white dwarfs. The one that accompanies Sirius, the brightest star in the heavens, has the mass of the sun but emits about 1/360 the light. Given their mass and size, white dwarfs must be humongously dense. Sirius's companion is some 61,000 times denser than water. What are these bizarre objects? Enter Sir Arthur Eddington.

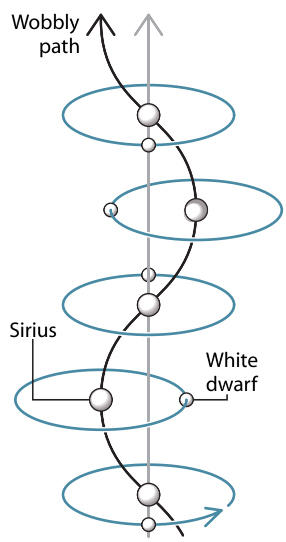
When I began studying physics in the late 1940s, Eddington was a hero of mine but for the wrong reasons. I knew nothing about his great work in astronomy. I admired his popular books (which, since I have learned more about physics, now seem rather silly to me). Eddington, who died in 1944, was a neo-Kantian who believed that everything of significance about the universe could be learned by examining what went on inside one's head. But starting in the late 1910s, when Eddington led one of the two expeditions that confirmed Einstein's prediction that the sun bends starlight, until the late 1930s, when Eddington really started going off the deep end, he was truly one of the giants of 20th-century science. He practically created the discipline that led to the first understanding of the internal constitution of stars, the title of his classic 1926 book. To him, white

An Early History of Black Holes



1900
Max Planck discovers black-body radiation.

1905
In a paper on black-body radiation, Albert Einstein shows that light can be viewed as particles (photons).



1915
Through spectroscopic studies, astronomer Walter S. Adams identifies Sirius's faint companion (which causes Sirius to wobble slightly as it moves) as a small, hot, dense star—a white dwarf.

1916
Einstein publishes his general theory of relativity, producing equations that describe gravity.

AIP EMILIO SEGRÈ VISUAL ARCHIVE (*Planck*); BETTMANN/CORBIS (*Einstein, 1905*); JARED SCHNEIDMAN DESIGN (*illustration*); dpa/CORBIS (*Einstein, 1916*)

THE AUTHOR

JEREMY BERNSTEIN is professor emeritus of physics at the Stevens Institute of Technology in Hoboken, N.J. He was a staff writer for the *New Yorker* from 1961 to 1995 and is the recipient of many science writing awards. He is a former adjunct professor at the Rockefeller University and a vice president of the board of trustees of the Aspen Center for Physics, of which he is now an honorary trustee. Bernstein has written 12 books on popular science and mountain travel. This article is adapted from his collection of essays, *A Theory for Everything,* published by Copernicus Books in 1996.

dwarfs were an affront, at least from an aesthetic point of view. But he studied them nonetheless and came up with a liberating idea.

In 1924 Eddington proposed that the gravitational pressure that was squeezing a dwarf might strip some of the electrons off protons. The atoms would then lose their "boundaries" and might be squeezed together into a small, dense package. The dwarf would eventually stop collapsing because of the Fermi-Dirac degeneracy pressure—that is, when the Pauli exclusion principle forced the electrons to recoil from one another.

The understanding of white dwarfs took another step forward in July 1930, when Subrahmanyan Chandrasekhar, who was 19, was onboard a ship sailing from Madras to Southampton. He had been accepted by British physicist R. H. Fowler to study with him at the University of Cambridge (where Eddington was, too). Having read Eddington's book on the stars and Fowler's book on quantum-statistical mechanics, Chandrasekhar had become fascinated by white dwarfs. To pass the time during the voyage, Chandrasekhar asked himself: Is there any upper limit to how massive a white dwarf can be before it collapses under the force of its own gravitation? His answer set off a revolution.

A white dwarf as a whole is electrically neutral, so all the electrons must have a corresponding proton, which is some 2,000 times more massive. Consequently, protons must supply the bulk of the gravitational compression. If the dwarf is not collapsing, the degeneracy pressure of the electrons and the gravitational collapse of the protons must just balance. This balance, it turns out, limits the number of protons and hence the mass of the dwarf. This maximum is known as the Chandrasekhar limit and equals about 1.4 times the mass of the sun. Any dwarf more massive than this number cannot be stable.

Chandrasekhar's result deeply disturbed Eddington. What happens if the mass is more than 1.4 times that of the sun? He was not pleased with the answer. Unless some mechanism could be found for limiting the mass of any star that was eventually going to compress itself into a dwarf, or unless Chandrasekhar's result was wrong, massive stars were fated to collapse gravitationally into oblivion.

Eddington found this intolerable and proceeded to attack Chandrasekhar's

1916
Karl Schwarzschild shows that a radius of a collapsing object exists at which Einstein's gravity equations become "singular"—time vanishes, and space becomes infinite.

1924
Einstein publishes Satyendra Nath Bose's work on black-body radiation, developing so-called quantum statistics for one class of particles (such as photons).

1924
Sir Arthur Eddington proposes that gravity can strip away electrons from protons in a white dwarf.

1925
Wolfgang Pauli formulates the exclusion principle, which states that certain particles cannot be in exactly the same quantum-mechanical state.

ROBERT BEIN *AIP Emilio Segrè Visual Archive* (*Schwarzschild*); MAX PLANCK INSTITUTE *Courtesy of AIP* (*Bose*); BETTMANN/CORBIS (*Eddington*); AIP EMILIO SEGRÈ VISUAL ARCHIVE (*Pauli*) use of quantum statistics—both publicly and privately. The criticism devastated Chandrasekhar. But he held his ground, bolstered by people such as Danish physicist Niels Bohr, who assured him that Eddington was simply wrong and should be ignored.

A Singular Sensation

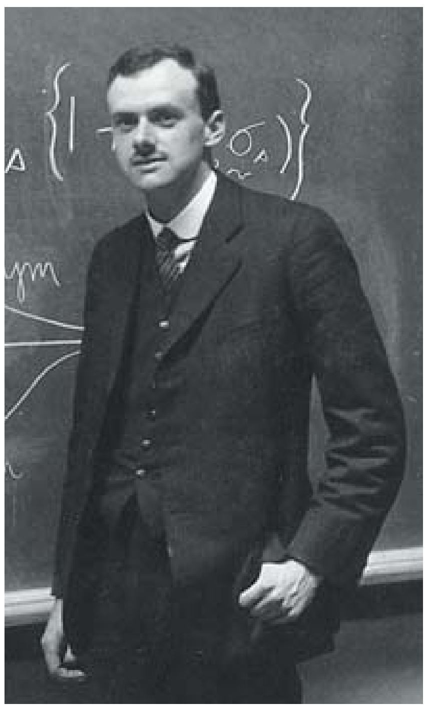
AS RESEARCHERS explored quantum statistics and white dwarfs, others tackled Einstein's work on gravitation, his general theory of relativity. As far as I know, Einstein never spent a great deal of time looking for exact solutions to his gravitational equations. The part that described gravity around matter was extremely complicated, because gravity distorts the geometry of space and time, causing a particle to move from point to point along a curved path. More important to Einstein, the source of gravity—matter—could not be described by the gravitational equations alone. It had to be put in by hand, leaving Einstein to feel the equations were incomplete. Still, approximate solutions could describe with sufficient accuracy phenomena such as the bending of starlight. Nevertheless, he was impressed when, in 1916, German astronomer Karl Schwarzschild came up with an exact solution for a realistic situation—in particular, the case of a planet orbiting a star.

Einstein felt that his gravitational equations were incomplete.

In the process, Schwarzschild found something disturbing. There is a distance from the center of the star at which the mathematics goes berserk. At this distance, now called the Schwarzschild radius, time vanishes, and space becomes infinite. The equation becomes what mathematicians call singular. The Schwarzschild radius is usually much smaller than the radius of the object. For the sun, for example, it is three kilometers, whereas for a one-gram marble it is 10^{-28} centimeter.

Schwarzschild was, of course, aware that his formula went crazy at this radius, but he decided that it did not matter. He constructed a simplified model of a star and showed that it would take an infinite gradient of pressure to compress it to his radius. The finding, he argued, served no practical interest.

But his analysis did not appease everybody. It bothered Einstein, because Schwarzschild's model star did not satisfy certain technical requirements of relativity theory. Various people, however, showed that one could rewrite Schwarzschild's solutions so that they avoided the singularity. But was the re-



1926
Enrico Fermi (*above left*) and P.A.M. Dirac (*above center*) develop quantum statistics for particles that obey Pauli's exclusion principle (such as electrons and protons). When compressed, such particles fly away from one another, creating a so-called degeneracy pressure.

1930
Using quantum statistics and Eddington's work on stars, Subrahmanyan Chandrasekhar finds that the upper mass limit for white dwarfs is 1.4 times the mass of the sun, suggesting that more massive stars collapse into oblivion. Eddington makes fun of him.

BETTMANN/CORBIS (*Fermi*); AIP EMILIO SEGRÈ VISUAL ARCHIVE (*Dirac*); BETTMANN/CORBIS (*Chandrasekhar*) sult really nonsingular? It would be incorrect to say that a debate raged, because most physicists had little regard for these matters—at least until 1939.

In his 1939 paper Einstein credits his renewed concern about the Schwarzschild radius to discussions with Princeton cosmologist Harold P. Robertson and with his assistant Peter G. Bergmann. It was certainly Einstein's intention in this paper to kill off the Schwarzschild singularity once and for all. At the end of it he writes, "The essential result of this investigation is a clear understanding as to why 'Schwarzschild singularities' do not exist in physical reality." In other words, black holes cannot exist.

To make his point, Einstein focused on a collection of small particles moving in circular orbits under the influence of one another's gravitation—in effect, a system resembling a spherical star cluster. He then asked whether such a configuration could collapse under its own gravity into a stable star with a radius equal to its Schwarzschild radius. He concluded that it could not, because at a somewhat larger radius the stars in the cluster would have to move faster than light in order to keep the configuration stable. Although Einstein's reasoning is correct, his point is irrelevant: it does not matter that a collapsing star at the Schwarzschild radius is unstable, because the star collapses past that radius anyway. I was much taken by the fact that the then 60-year-old Einstein presents in this paper tables of numerical results, which he must have gotten by using a slide rule. But the paper, like the slide rule, is now a historical artifact.

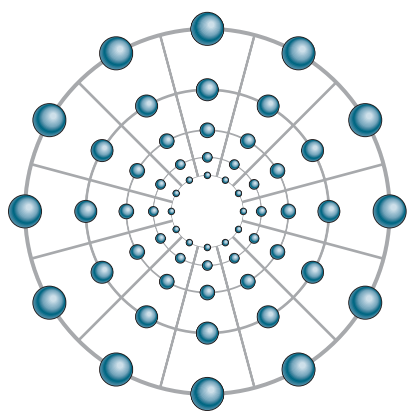
From Neutrons to Black Holes

WHILE EINSTEIN was doing this research, an entirely different enterprise was unfolding in California. Oppenheimer and his students were creating the modern theory of black holes. The curious thing about the black hole research is that it was inspired by an idea that turned out to be entirely wrong. In 1932 British experimental physicist James Chadwick found the neutron, the neutral component of the atomic nucleus. Soon thereafter speculation began—most notably by Fritz Zwicky of the California Institute of Technology and independently by the brilliant Soviet theoretical physicist Lev D. Landau—that neutrons could lead to an alternative to white dwarfs.

When the gravitational pressure got large enough, they argued, an electron in a star could react with a proton to produce a neutron. (Zwicky even conjectured that this process would happen in supernova explosions; he was right, and these "neutron stars" we now identify as pulsars.) At the time of this work, the actual mechanism for generating the energy in ordinary stars was not known. One solution placed a neutron star at the center of ordinary stars, in somewhat the same spirit that

1932
James Chadwick discovers the neutron. Its existence leads researchers to wonder if "neutron stars" could be an alternative to white dwarfs.

"On a Stationary System with Spherical Symmetry Consisting of Many Gravitating Masses"
—*Albert Einstein,* Annals of Mathematics, *1939*



1939
Sparked by conversations with colleagues, Einstein tries to kill off the Schwarzschild radius once and for all: he concludes that black holes are impossible in a paper published in the *Annals of Mathematics.*

1939
Using ideas of collapsing neutron stars and white dwarfs, J. Robert Oppenheimer and his student Hartland S. Snyder show how a black hole can form.

BETTMANN/CORBIS (*Chadwick*); JARED SCHNEIDMAN DESIGN (*star cluster illustration*); BETTMANN/CORBIS (*Oppenheimer*) many astrophysicists now conjecture that black holes power quasars.

The question then arose: What was the equivalent of the Chandrasekhar mass limit for these stars? Determining this answer is much harder than finding the limit for white dwarfs. The reason is that the neutrons interact with one another with a strong force whose specifics we still do not fully understand. Gravity will eventually overcome this force, but the precise limiting mass is sensitive to the details. Oppenheimer published two papers on this subject with his students Robert Serber and George M. Volkoff and concluded that the mass limit here is comparable to the Chandrasekhar limit for white dwarfs. The first of these papers was published in 1938 and the second in 1939. (The real source of stellar energy—fusion—was discovered in 1938 by Hans Bethe and Carl Friedrich von Weizsäcker, but it took a few years to be accepted, and so astrophysicists continued to pursue alternative theories.)

Oppenheimer went on to ask exactly what Eddington had wondered about white dwarfs: What would happen if one had a collapsing star whose mass exceeded any of the limits? Einstein's 1939 rejection of black holes—to which Oppenheimer and his students were certainly oblivious, for they were working concurrently, 3,000 miles away—was of no relevance. But Oppenheimer did not want to construct a stable star with a radius equal to its Schwarzschild radius. He wanted to see what would happen if one let the star collapse through its Schwarzschild radius. He suggested that Snyder work out this problem in detail.

To simplify matters, Oppenheimer told Snyder to make certain assumptions and to neglect technical considerations such as the degeneracy pressure or the possible rotation of the star. Oppenheimer's intuition told him that these factors would not change anything essential. (These assumptions were challenged many years later by a new generation of researchers using sophisticated high-speed computers—poor Snyder had an old-fashioned mechanical desk

> Although Einstein's reasoning is correct, **his point is irrelevant.**

calculator—but Oppenheimer was right. Nothing essential changes.) With the simplified assumptions, Snyder found out that what happens to a collapsing star depends dramatically on the vantage point of the observer.

Two Views of a Collapse

LET US START with an observer at rest a safe distance from the star. Let us also suppose that there is another observer attached to the surface of the star—"co-moving" with its collapse—who can send light signals back to his stationary colleague. The stationary observer will see the signals from his moving counterpart gradually shift to the red end of the electromagnetic spectrum. If the frequency of the signals is thought of as a clock, the stationary observer will say that the moving observer's clock is gradually slowing down.

Indeed, at the Schwarzschild radius the clock will slow down to zero. The stationary observer will argue that it took an infinite amount of time for the star to collapse to its Schwarzschild radius. What happens after that we cannot say, because, according to the stationary observer, there is no "after." As far as this observer is concerned, the star is frozen at its Schwarzschild radius.

Indeed, until December 1967, when physicist John A. Wheeler of Princeton University coined the name "black hole" in a lecture he presented, these objects were often referred to in the literature as frozen stars. This frozen state is the real significance of the singularity in the Schwarzschild geometry. As Oppenheimer and Snyder observed in their paper, the collapsing star "tends to close itself off from any communication with a distant observer; only its gravitational field persists." In other words, a black hole has been formed.

But what about observers riding with collapsing stars? These observers, Oppenheimer and Snyder pointed out, have a completely different sense of things. To them, the Schwarzschild radius has no special significance. They pass right through it and on to the center in a matter of hours, as measured by their watches. They would, however, be subject to monstrous tidal gravitational forces that would tear them to pieces.

The year was 1939, and the world itself was about to be torn to pieces. Oppenheimer was soon to go off to war to build the most destructive weapon ever devised by humans. He never worked on the subject of black holes again. As far as I know, Einstein never did, either. In peacetime, in 1947, Oppenheimer became the director of the Institute for Advanced Study in Princeton, N.J., where Einstein was a professor. From time to time they talked. There is no record of their ever having discussed black holes. Further progress would have to wait until the 1960s, when discoveries of quasars, pulsars and compact x-ray sources reinvigorated thinking about the mysterious fate of stars. SA

MORE TO EXPLORE

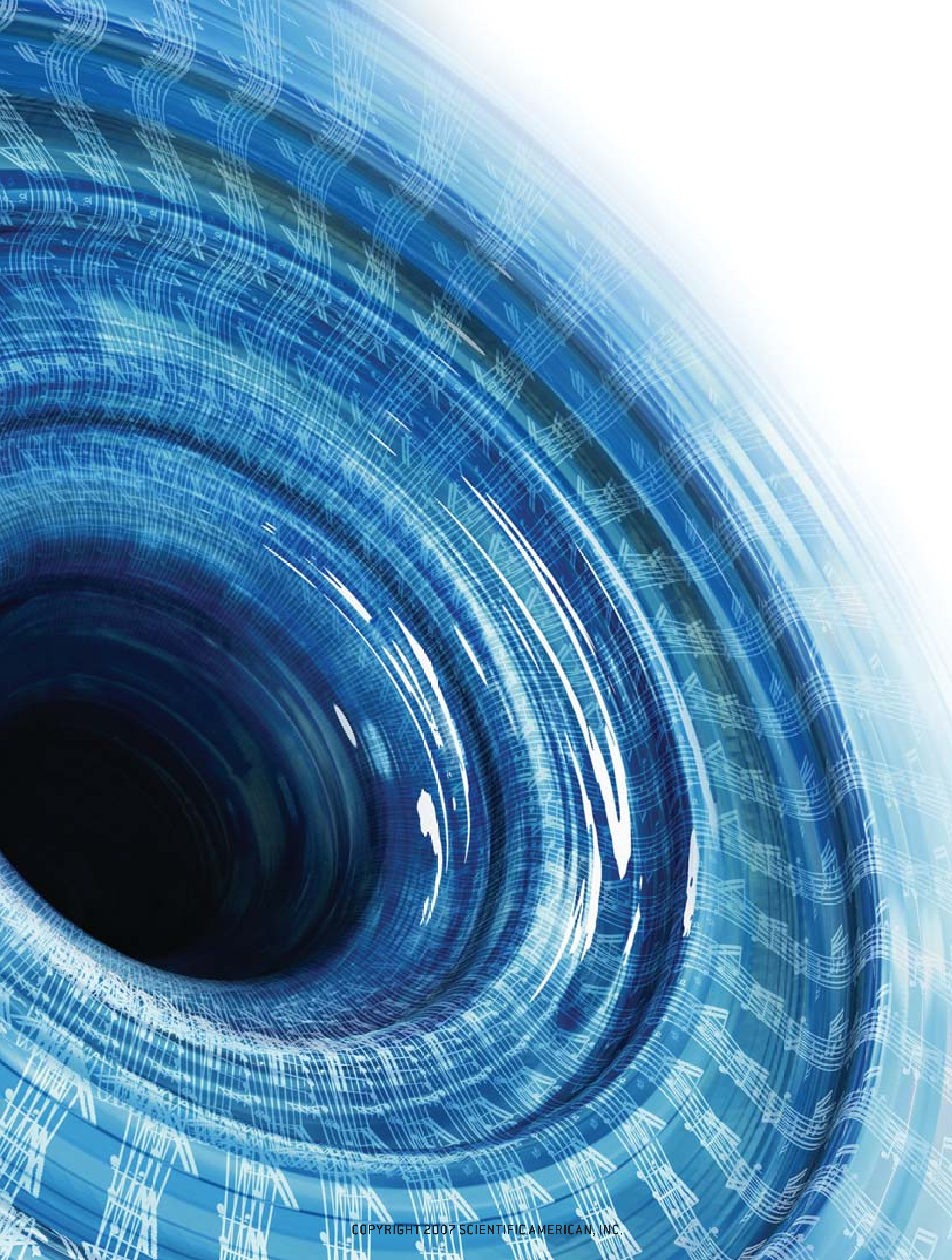
Subtle Is the Lord: The Science and the Life of Albert Einstein. Abraham Pais. Oxford University Press, 1982.

Dark Stars: The Evolution of an Idea. Werner Israel in *300 Years of Gravitation.* Edited by S. W. Hawking and W. Israel. Cambridge University Press, 1987.

Chandra: A Biography of S. Chandrasekhar. Kameshwar C. Wali. University of Chicago Press, 1991.

Black Holes. J.-P. Luminet et al. Cambridge University Press, 1992.

Black Holes and Time Warps. Kip Thorne. W. W. Norton, 1994.



AN ECHO of black holes

Sound waves in a fluid behave uncannily like light waves in space. Black holes even have acoustic counterparts. Could spacetime literally be a kind of fluid, like the ether of pre-Einsteinian physics?

By Theodore A. Jacobson and Renaud Parentani

When Albert Einstein proposed his special theory of relativity in 1905, he rejected the 19th-century idea that light arises from vibrations of a hypothetical medium, the "ether." Instead, he argued, light waves can travel in vacuo without being supported by any material—unlike sound waves, which are vibrations of the medium in which they propagate. This feature of special relativity is untouched in the two other pillars of modern physics, general relativity and quantum mechanics. Right up to the present day, all experimental data, on scales ranging from subnuclear to galactic, are successfully explained by these three theories.

Nevertheless, physicists face a deep conceptual problem. As currently understood, general relativity and quantum mechanics are incompatible. Gravity, which general relativity attributes to the curvature of the spacetime continuum, stubbornly resists being incorporated into a quantum framework. Theorists have made only incremental progress toward understanding the highly curved structure of spacetime that quantum mechanics leads them to expect at extremely short distances. Frustrated, some have turned to an unexpected source for guidance: condensed-matter physics, the study of common substances such as crystals and fluids.

Like spacetime, condensed matter looks like a continuum when viewed at large scales, but unlike spacetime it has a well-understood microscopic structure governed by quantum mechanics. Moreover, the propagation of sound in an uneven fluid flow is closely analogous to the propagation of light in a curved spacetime. By studying a model of a black hole using sound waves, we and our colleagues are attempting to exploit this analogy to gain insight into the possible microscopic workings of spacetime. The work suggests that spacetime may, like a material fluid, be granular and possess a preferred frame of reference that manifests itself on fine scales—contrary to Einstein's assumptions.

KENN BROWN

From Black Hole to Hot Coal

BLACK HOLES are a favorite testing ground for quantum gravity because they are among the few places where quantum mechanics and general relativity are both critically important. A major step toward a merger of the two theories came in 1974, when Stephen Hawking of the University of Cambridge applied quantum mechanics to the horizon of black holes.

According to general relativity, the horizon is the surface that separates the inside of a black hole (where gravity is so strong that nothing can escape) from the outside. It is not a material limit; unfortunate travelers falling into the hole would not sense anything special on crossing the horizon. But once having done so, they would no longer be able to send light signals to people outside, let alone return there. An outside observer would receive only the signals transmitted by the travelers before they crossed over. As light waves climb out of the gravitational well around a black hole, they get stretched out, shifting down in frequency and lengthening in duration. Consequently, to the observer, the travelers would appear to move in slow motion and to be redder than usual.

This effect, known as gravitational redshift, is not specific to black holes. It also alters the frequency and timing of signals between, say, orbiting satellites and ground stations. GPS navigation systems must take it into account to work accurately. What is specific to black holes, however, is that the redshift becomes infinite as the travelers approach the horizon. From the outside observer's point of view, the descent appears to take an infinite amount of time, even though only a finite time passes for the travelers themselves.

So far this description of black holes has treated light as a classical electromagnetic wave. What Hawking did was to reconsider the implications of the infinite redshift when the quantum nature of light is taken into account. According to quantum theory, even a perfect vacuum is not truly empty;

RIPPLES IN A STREAM behave much like light waves in spacetime. The flow of the stream around the rock is not uniform, so the ripples are bent and their wavelengths vary. The same happens to light passing through the gravitational field of a planet or star. In some cases, the flow is so fast that ripples cannot propagate upstream—just as light cannot propagate out of a black hole.

BOB FOSBURY

it is filled with fluctuations as a result of the Heisenberg uncertainty principle. The fluctuations take the form of pairs of virtual photons. These photons are called virtual because, in an uncurved spacetime, far from any gravitational influence, they appear and disappear restlessly, remaining unobservable in the absence of any disturbance.

But in the curved spacetime around a black hole, one member of the pair can be trapped inside the horizon while the other gets stranded outside. The pair can then pass from virtual to real, leading to an outward flux of observable light and a corresponding decrease in the mass of the hole. The overall pattern of radiation is thermal, like that from a hot coal, with a temperature inversely proportional to the mass of the black hole. This phenomenon is called the Hawking effect. Unless the hole swallows matter or energy to make up the loss, the Hawking radiation will drain it of all its mass.

An important point—which will become critical later when considering fluid analogies to black holes—is that the space very near the black hole horizon remains a nearly perfect quantum vacuum. In fact, this condition is essential for Hawking's argument. The virtual photons are a feature of the lowest-energy quantum state, or "ground state." It is only in the process of separating from their partners and climbing away from the horizon that the virtual photons become real.

The Ultimate Microscope

HAWKING'S ANALYSIS has played a central role in the attempt to build a full quantum theory of gravity. The ability to reproduce and elucidate the effect is a crucial test for candidate quantum gravity theories, such as string theory. Yet although most physicists accept Hawking's argument, they have never been able to confirm it experimentally. The predicted emission from stellar and galactic black holes is far too feeble to see. The only hope for observing Hawking radiation is to find miniature holes left over from the early universe or created in particle accelerators, which may well prove impossible.

The lack of empirical confirmation of the Hawking effect is particularly vexing in view of the disturbing fact that the theory has potential flaws, stemming from the infinite redshift that it predicts a photon will undergo. Consider what the emission process looks like when viewed reversed in time. As the Hawking photon gets nearer to the hole, it blueshifts to a higher frequency and correspondingly shorter wavelength. The further back in time it is followed, the closer it approaches the horizon and the shorter its wavelength becomes. Once the wavelength becomes much smaller than the black hole, the particle joins its partner and becomes the virtual pair discussed earlier.

The blueshifting continues without abatement, down to arbitrarily short distances. Smaller than a distance of about 10^{-35} meter, known as the Planck length, neither relativity nor standard quantum theory can predict what the particle will do. A quantum theory of gravity is needed. A black hole horizon thus acts as a fantastic microscope that brings the observer into contact with unknown physics. For a theorist, this magnification is worrisome. If Hawking's prediction relies on unknown physics, should we not be suspicious of its validity? Might the properties, even

Overview
Acoustic Black Holes

- The famous physicist Stephen Hawking argued in the 1970s that black holes are not truly black; they emit a quantum glow of thermal radiation. But his analysis had a problem. According to relativity theory, waves starting at a black hole horizon will be stretched by an infinite amount as they propagate away. Therefore, Hawking's radiation must emerge from an infinitely small region of space, where the unknown effects of quantum gravity take over.
- Physicists have grappled with this problem by studying black hole analogues in fluid systems. The fluid's molecular structure cuts off the infinite stretching and replaces the microscopic mysteries of spacetime by known physics.
- The analogies lend credence to Hawking's conclusion. They also suggest to some researchers that spacetime has a "molecular" structure, contrary to the assumptions of standard relativity theory.

the existence, of Hawking radiation depend on the microscopic properties of spacetime—much as, for example, the heat capacity or speed of sound of a substance depends on its microscopic structure and dynamics? Or is the effect, as Hawking originally argued, entirely determined just by the macroscopic properties of the black hole, namely, its mass and spin?

Sound Bites

ONE EFFORT TO ANSWER these embarrassing questions began with the work of William Unruh of the University of British Columbia. In 1981 he showed that there is a close analogy between the propagation of sound in a moving fluid and that of light in a curved spacetime. He suggested that this analogy might be useful in assessing the impact of microscopic physics on the origin of Hawking radiation. Moreover, it might even allow for experimental observation of a Hawking-like phenomenon.

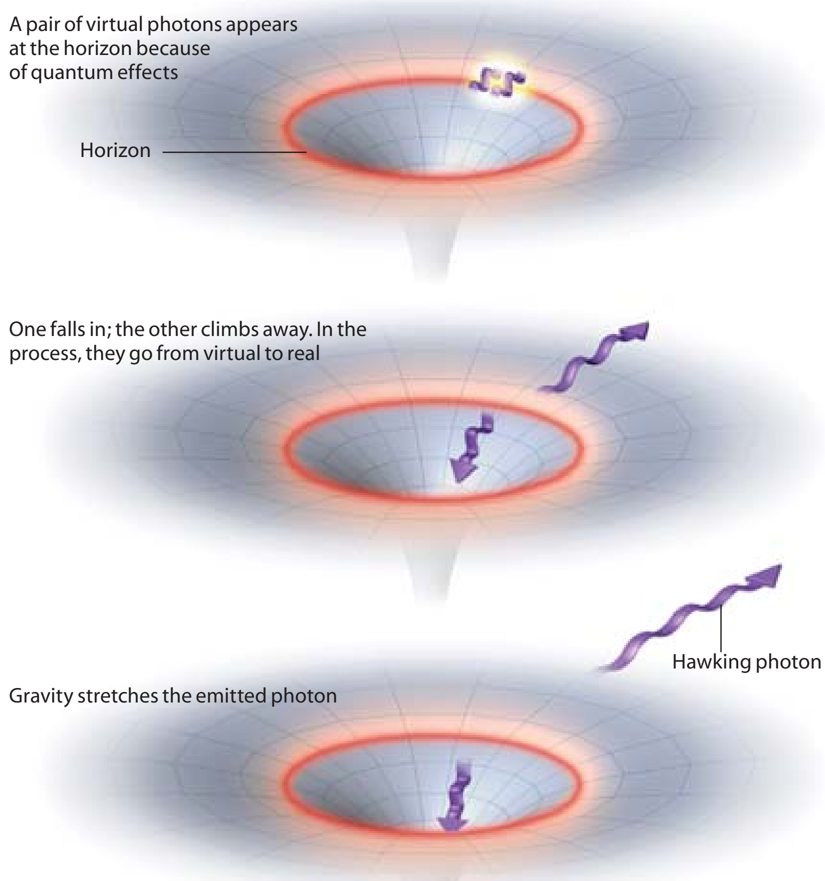
Like light waves, acoustic (sound) waves are characterized by a frequency, wavelength and propagation speed. The very concept of a sound wave is valid only when the wavelength is much longer than the distance between molecules of the fluid; on smaller scales, acoustic waves cease to exist. It is precisely this limitation that makes the analogy so interesting, because it can allow physicists to study the macroscopic consequences of microscopic structure. To be truly useful, however, this analogy must extend to the quantum level. Ordinarily, random thermal jigging of the molecules prevents sound waves from behaving analogously to light quanta. But when the temperature approaches absolute zero, sound can behave like quantum particles, which physicists call "phonons" to underline the analogy with the particles of light, photons. Experimenters routinely observe phonons in crystals and in substances that remain fluid at sufficiently low temperatures, such as liquid helium.

The behavior of phonons in a fluid at rest or moving uniformly is like that of photons in flat spacetime, where gravity is absent. Such phonons propa-

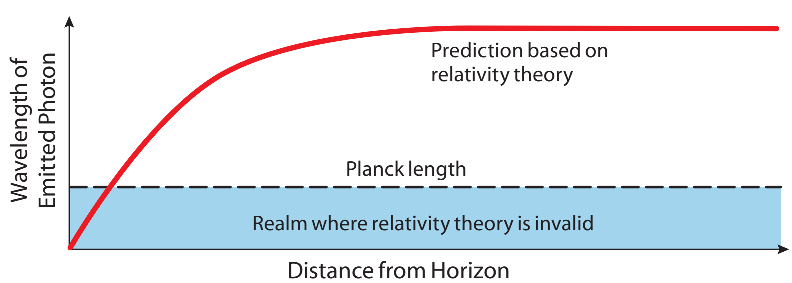
GEORGE RETSECK

Was Hawking Wrong?

One of the greatest—and least recognized—mysteries of black holes concerns a flaw in Stephen Hawking's famous prediction that black holes emit radiation. A hole is defined by an event horizon, a one-way door: objects on the outside can fall in, but objects on the inside cannot get out. Hawking asked what happens to pairs of virtual particles (which continually appear and disappear everywhere in empty space because of quantum effects) that originate at the horizon itself.



Relativity theory predicts that a photon from the horizon gets stretched by an infinite amount (*red curve, below*). In other words, an observed photon must have originated as a virtual one with a wavelength of almost precisely zero, which is problematic because unknown quantum gravity effects take over at distances shorter than the so-called Planck length of 10^{-35} meter. This conundrum has driven physicists to design experimentally realizable analogues to black holes to see whether they indeed emit radiation and to understand how it originates.



Light vs. Sound

Type of Wave	Classical Description	Quantum Description	Velocity	What Causes Path of Wave to Curve	Where Description Breaks Down
Light	Oscillating electric and magnetic fields	Electromagnetic-wave photon	300,000 kilometers per second	Spacetime curvature, caused by matter and energy	Planck length? (10^{-35} meter)
Sound	Collective movements of molecules	Acoustic-wave phonon	1,500 meters per second (in liquid water)	Variations in fluid speed and direction	Intermolecular distance (10^{-10} meter for water)

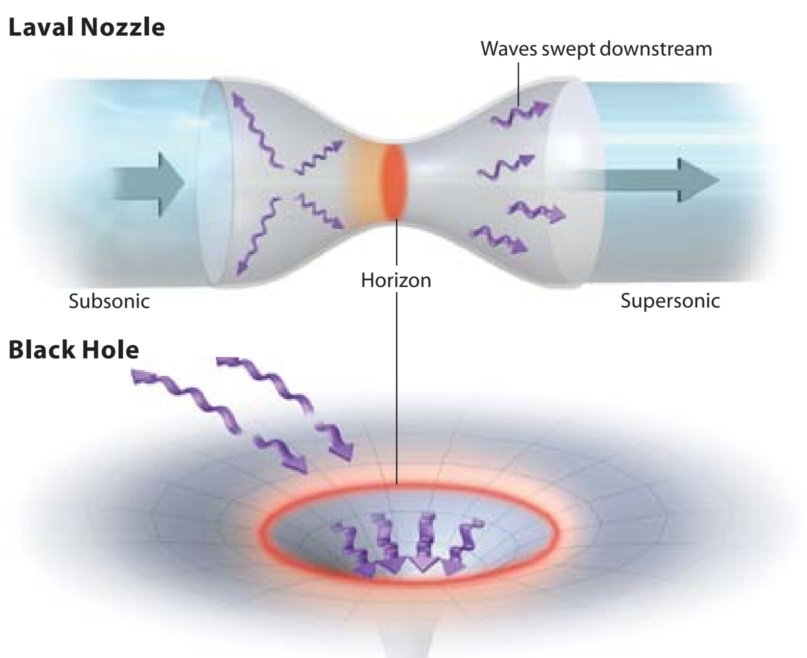
gate in straight lines with unchanging wavelength, frequency and velocity. Sound in, say, a swimming pool or a smoothly flowing river travels straight from its source to the ear.

In a fluid moving nonuniformly, however, the phonons' velocity is altered and their wavelength can become stretched, just like photons in a curved spacetime. Sound in a river entering a narrow canyon or water swirling down the drain becomes distorted and follows a bent path, like light around a star. In fact, the situation can be described using the geometric tools of general relativity.

A fluid flow can even act on sound as a black hole acts on light. One way to create such an acoustic black hole is to use a device that hydrodynamicists call a Laval nozzle. The nozzle is designed so that the fluid reaches the speed of sound at the narrowest point and is supersonic beyond it. The effective acoustic geometry is very similar to the spacetime geometry of a black hole. The supersonic region corresponds to the hole's interior: sound waves propagating against the direction of the flow are swept downstream, like light pulled toward the center of a hole. The subsonic region is the exterior of the hole: sound waves can propagate upstream but only at the expense of being stretched, like light being redshifted. The boundary between the two regions behaves exactly like a black hole horizon.

Black Hole Analogue

A Laval nozzle—found at the end of rockets—makes a ready analogue to a black hole. The incoming fluid is subsonic; the constriction forces it to accelerate to the speed of sound, so that the outgoing fluid is supersonic. Sound waves in the subsonic region can move upstream, whereas waves in the supersonic region cannot. The constriction thus acts just like the horizon of a black hole: sound can enter but not exit the supersonic region. Quantum fluctuations in the constriction should generate sound analogous to Hawking radiation.



Atomism

IF THE FLUID is cold enough, the analogy extends to the quantum level. Unruh argued that the sonic horizon emits thermal phonons analogous to Hawking radiation. Quantum fluctuations near the horizon cause pairs of phonons to appear; one partner gets swept into the supersonic region, never to return, while the other ripples upstream, getting stretched out by the fluid flow. A microphone placed upstream picks up a faint hiss. The sound energy of the hiss is drawn from the kinetic energy of the fluid flow.

The dominant tone of the noise depends on the geometry; the typical wavelength of the observed phonons is comparable to the distance over which the flow velocity changes appreciably. This distance is much larger than the distance between molecules, so Unruh did his original analysis assuming that the fluid is smooth and continuous. Yet the phonons originate near the horizon with wavelengths so short that they should be sensitive to the granularity of the fluid. Does that affect the end result? Does a real fluid emit Hawking-like phonons, or is Unruh's prediction an artifact of the idealization of a continuous fluid? If that question can be answered for acoustic black holes, it may by analogy guide physicists in the case of gravitational black holes.

Physicists have proposed a number of black hole analogues besides the transsonic fluid flow. One involves not sound waves but ripples on the surface of a liquid or along the interface between layers of superfluid helium, which is so cold that it has lost all frictional resistance to motion. Recently Unruh and Ralf Schützhold of the Technical University of Dresden in Germany proposed to study electromagnetic waves passing through a tiny, carefully engineered electronic pipe. By sweeping a laser along the pipe to change the local wave speed, physicists might be able to create a horizon. Yet another idea is to model the accelerating expansion of the universe, which generates a Hawking-like radiation. A Bose-Einstein condensate—a gas so cold that the atoms have lost their individual identity—can act on sound like an expanding universe does on light, either by literally flying apart or by being manipulated using a magnetic field to give the same effect.

As yet, experimenters have not created any of these devices in the laboratory. The procedures are complicated, and experimenters have plenty of other low-temperature phenomena to keep them busy. So theorists have been working to see whether they can make headway on the problem mathematically.

Understanding how the molecular structure of the fluid affects phonons is extremely complicated. Fortunately, 10 years after Unruh proposed his sonic analogy, one of us (Jacobson) came up with a very useful simplification. The essential details of the molecular structure are encapsulated in the way that the frequency of a sound wave depends on its wavelength. This dependence, called the dispersion relation, determines the velocity of propagation. For large wavelengths, the velocity remains constant. For short wavelengths, approaching the intermolecular distance, the velocity can vary with wavelength.

Three different behaviors can arise. Type I is no dispersion—the wave behaves the same at short wavelengths as it does at long ones. For type II, the velocity decreases as the wavelength decreases, and for type III, velocity increases. Type I describes photons in relativity. Type II describes phonons in, for example, superfluid helium, and type III describes phonons in dilute Bose-Einstein condensates. This division into three types provides an organizing principle for figuring out how molecular structure affects sound on a macroscopic level. Beginning in 1995, Unruh and then other researchers have examined the Hawking effect in the presence of type II and type III dispersion.

Consider how the Hawking-like phonons look when viewed backward in time. Initially the dispersion type does not matter. The phonons swim downstream toward the horizon, their wavelengths decreasing all the while. Once the wavelength approaches the intermolecular distance, the specific dispersion relation becomes important. For type II, the phonons slow down, then reverse direction and start heading upstream again. For type III, they accelerate, break the long-wavelength speed of sound, then cross the horizon.

Ether Redux

A TRUE ANALOGY to the Hawking effect must meet an important condition: the virtual phonon pairs must begin life in their ground state, as do the virtual photon pairs around the black hole. In a real fluid, this condition would be easily met. As long as the macroscopic fluid flow changes slowly in time and space (compared with the pace of events at the molecular level), the molecular state continuously adjusts to minimize the energy of the system as a whole. It does not matter which molecules the fluid is made of.

With this condition met, it turns out that the fluid emits Hawking-like radiation no matter which of the three types

Energy Balance in the Hawking Effect

A common source of confusion in understanding the Hawking effect is how the energy balance is accounted for in the process and what is happening with the "virtual pairs" that are the origin of the radiation. Consider a pair of photons emerging from the vacuum, one outside the horizon with positive energy and the other inside with opposite, negative energy. (The members of a virtual pair must always have opposite values of energy, because the total energy is conserved.) Negative-energy particles cannot exist outside the horizon, because the vacuum is by definition the lowest energy state. Therefore, only a positive-energy photon can escape, whereas its negative-energy partner is trapped inside, lowering the total energy—and therefore the mass—of the black hole.

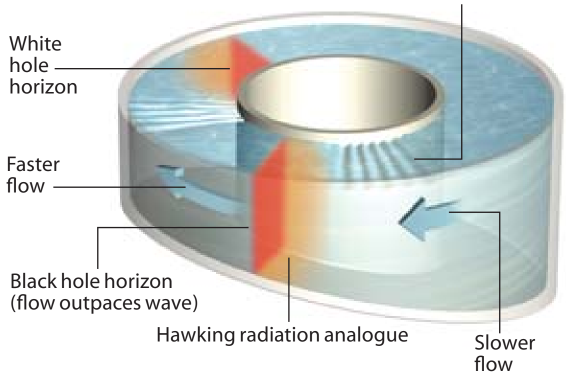
If a negative-energy photon cannot exist outside the horizon, how can it exist inside? Would not that violate the definition of the vacuum, too? To understand why not, we must distinguish between locally measured energy and globally conserved energy. The usual concept of conserved energy is related to time-shift symmetry, whereby the laws of physics are the same at all times. Conserved momentum is related to space-shift symmetry. In a black hole spacetime, the global symmetry that is a temporal shift outside the horizon becomes a spatial shift inside. So the single conserved quantity, the "global energy," corresponds to energy outside and momentum inside. In the Hawking effect, the partner photons inside the horizon have negative "global energy," but their locally measured energy is positive.

In the fluid analogue of a black hole, the energy for the sonic Hawking radiation comes from the kinetic energy of the bulk flow of fluid. A sound wave going upstream saps energy from the flow, but the energy of the wave itself makes up for this, so the total energy is higher—as long as the flow speed is less than the speed of sound. Inside the sonic horizon, the flow speed is greater than the speed of sound. There the wave saps more energy from the flow than it carries itself, so the total energy is less than that of the undisturbed flow. Such a wave can be thought of as containing negative energy. *—T.A.J. and R.P.*

Other Black Hole Models

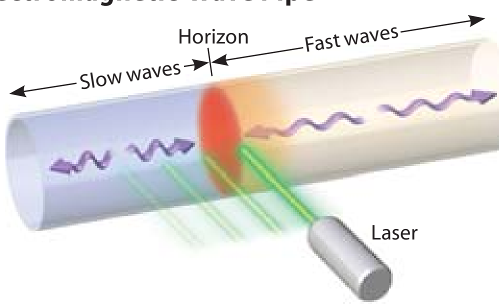
Devices besides the Laval nozzle also reproduce the essential characteristic of a black hole horizon: waves can go one way but not the other. Each offers novel insights into black holes. All should generate the analogue of Hawking radiation.

Surface Ripples



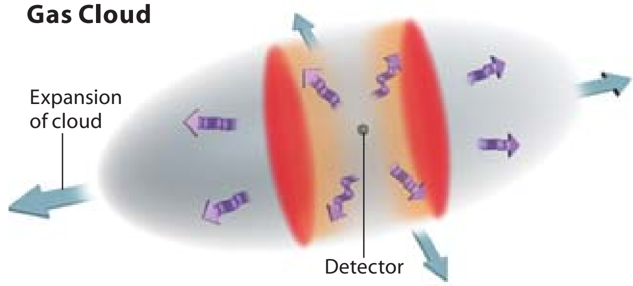
Instead of sound waves, this experiment involves surface waves in liquid flowing around a circular channel. As the channel becomes shallower, the flow speeds up and, at some point, outpaces the waves, preventing them from traveling upstream—thereby creating the analogue of a black hole horizon. Completing the circuit is the horizon of a "white hole": a body that lets material flow out but not in. To observe Hawking-like radiation would require a supercooled fluid such as helium 4.

Electromagnetic-Wave Pipe



This experiment studies microwaves passing through a rod built so that the speed of wave propagation can be tweaked with a laser beam. Sweeping the beam along the rod creates a moving horizon that divides the rod into slow- and fast-wave zones. Waves in the slow zone cannot reach the fast zone, but waves in the fast zone can cross to the slow. The Hawking-like radiation may be stronger and easier to observe than in fluid analogies.

Gas Cloud



The long axis of an inflating, cigar-shaped gas cloud can simulate a one-dimensional universe expanding at an accelerating rate. Such a universe behaves like an inside-out black hole: waves outside the horizons are swept away too quickly to enter the inner region. A Hawking-like radiation should stream inward. In practice, the gas would be a Bose-Einstein condensate, a supercooled gas with quantum properties that make the Hawking analogy possible.

of dispersion relations applies. The microscopic details of the fluid do not have any effect. They get washed out as the phonons travel away from the horizon. In addition, the arbitrarily short wavelengths invoked by original Hawking analysis do not arise when either type II or III dispersion is included. Instead the wavelengths bottom out at the intermolecular distance. The infinite redshift is an avatar of the unphysical assumption of infinitely small atoms.

Applied to real black holes, the fluid analogy lends confidence that Hawking's result is correct despite the simplifications he made. Moreover, it suggests to some researchers that the infinite redshift at a gravitational black hole horizon may be similarly avoided by dispersion of short wavelength light. But there is a catch. Relativity theory flatly asserts that light does not undergo dispersion in a vacuum. The wavelength of a photon appears different to different observers; it is arbitrarily long when viewed from a reference frame that is moving sufficiently close to the speed of light. Hence, the laws of physics cannot mandate a fixed short-wavelength cutoff, at which the dispersion relation changes from type I to type II or III. Each observer would perceive a different cutoff.

THE AUTHORS

THEODORE A. JACOBSON and *RENAUD PARENTANI* study the puzzles of quantum gravity and its possible observable consequences for black holes and cosmology. Jacobson is a physics professor at the University of Maryland. His recent research focuses on the thermodynamics of black holes, how spacetime might be microscopically discrete and whether that fine structure could be macroscopically detected. Parentani is a physics professor at the University of Paris-Sud at Orsay who does research at the CNRS Laboratory of Theoretical Physics. He investigates the role of quantum fluctuations in black hole physics and cosmology.

GEORGE RETSECK

Physicists thus face a dilemma. Either they retain Einstein's injunction against a preferred frame and they swallow the infinite redshifting, or they assume that photons do not undergo an infinite redshift and they have to introduce a preferred reference frame. Would this frame necessarily violate relativity? No one yet knows. Perhaps the preferred frame is a local effect that arises only near black hole horizons—in which case relativity continues to apply in general. On the other hand, perhaps the preferred frame exists everywhere, not just near black holes—in which case relativity is merely an approximation to a deeper theory of nature. Experimenters have yet to see such a frame, but the null result may simply be for want of sufficient precision.

Physicists have long suspected that reconciling general relativity with quantum mechanics would involve a short-distance cutoff, probably related to the Planck scale. The acoustic analogy bolsters this suspicion. Spacetime must be somehow granular to tame the dubious infinite redshift.

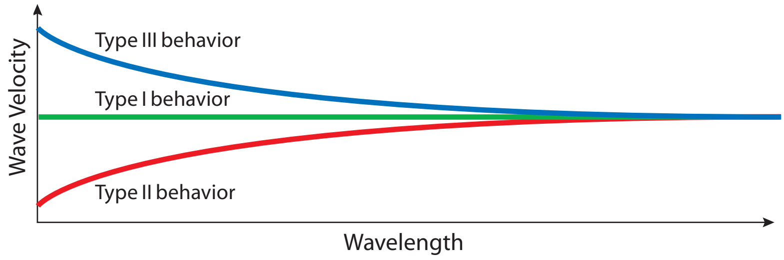
If so, the analogy between sound and light propagation would be even better than Unruh originally thought. The unification of general relativity and quantum mechanics may lead us to abandon the idealization of continuous space and time and to discover the "atoms" of spacetime. Einstein may have had similar thoughts when he wrote to his close friend Michele Besso in 1954, the year before his death: "I consider it quite possible that physics cannot be based on the field concept, that is, on continuous structures." But this would knock out the very foundation from under physics, and at present scientists have no clear candidate for a substitute. Indeed, Einstein went on to say in his next sentence, "Then *nothing* remains of my entire castle in the air, including the theory of gravitation, but also nothing of the rest of modern physics."

Fifty years later the castle remains intact, although its future is unclear. Black holes and their acoustic analogues have perhaps begun to light the path and sound out the way. SA

GEORGE RETSECK

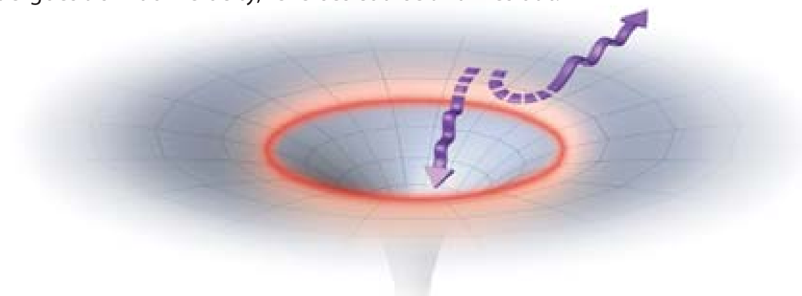
Hawking Was Right, but ...

The fluid analogies suggest how to fix Hawking's analysis. In an idealized fluid, the speed of sound is the same no matter the wavelength (so-called type I behavior). In a real fluid, the speed of sound either decreases (type II) or increases (type III) as the wavelength approaches the distance between molecules.

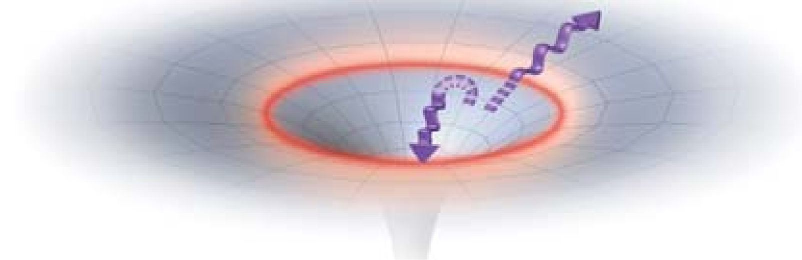


Hawking's analysis is based on standard relativity theory, in which light travels at a constant speed—type I behavior. If its speed varied with wavelength, as in the fluid analogues, the paths of the Hawking photons would change.

For type II, the photons originate outside the horizon and fall inward. One undergoes a shift of velocity, reverses course and flies out.



For type III, the photons originate inside the horizon. One accelerates past the usual speed of light, allowing it to escape.



Because the photons do not originate exactly at the horizon, they do not become infinitely redshifted. This fix to Hawking's analysis has a price: relativity theory must be modified. Contrary to Einstein's assumptions, spacetime must act like a fluid consisting of some unknown kind of "molecules."

MORE TO EXPLORE

Trans-Planckian Redshifts and the Substance of the Space-Time River. Ted Jacobson in *Progress of Theoretical Physics Supplement*, No. 136, pages 1–17; 1999. Available online at **http://ptp.ipap.jp/cgi-bin/getarticle?magazine=PTPS&volume=136&number=&page=1-17**

What Did We Learn from Studying Acoustic Black Holes? Renaud Parentani in *International Journal of Modern Physics A*, Vol. 17, No. 20, pages 2721–2726; August 10, 2002. **arxiv.org/abs/gr-qc/0204079**

Black-Hole Physics in an Electromagnetic Waveguide. Steven K. Blau in *Physics Today*, Vol. 58, No. 8, pages 19–20; August 2005.

Analogue Gravity. Carlos Barceló, Stefano Liberati and Matt Visser in *Living Reviews in Relativity*, Vol. 8, No. 12; 2005. Available at **www.livingreviews.org/lrr-2005-12**

QUANTUM black holes

Physicists could soon be creating black holes in the laboratory

By Bernard J. Carr and Steven B. Giddings

Ever since physicists invented particle accelerators, nearly 80 years ago, they have used them for such exotic tasks as splitting atoms, transmuting elements, producing antimatter and creating particles not previously observed in nature. With luck, though, they could soon undertake a challenge that will make those achievements seem almost pedestrian. Accelerators may produce the most profoundly mysterious objects in the universe: black holes.

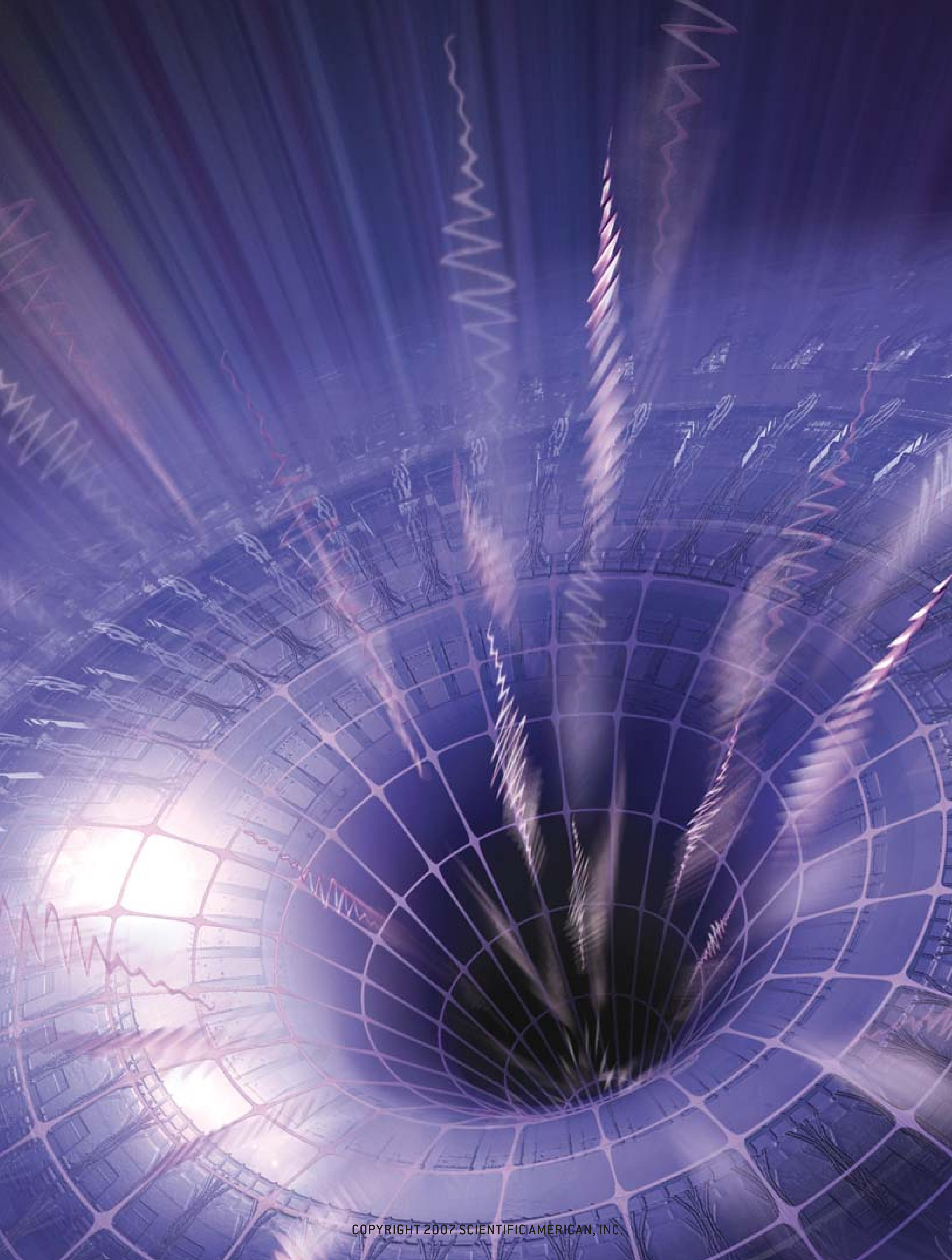
When one thinks of black holes, one usually envisions massive monsters that can swallow spaceships, or even stars, whole. But the holes that might be produced at the highest-energy accelerators—perhaps as early as mid-2008, when the Large Hadron Collider (LHC) at CERN near Geneva starts running at full design energy—are distant cousins of such astrophysical behemoths. They would be microscopic, comparable in size to elementary particles. They would not rip apart stars, reign over galaxies or pose a threat to our planet, but in some respects their properties should be even more dramatic. Because of quantum effects, they would evaporate shortly after they formed, lighting up the particle detectors like Christmas trees. In so doing, they could give clues about how spacetime is woven together and whether it has unseen higher dimensions.

A Tight Squeeze

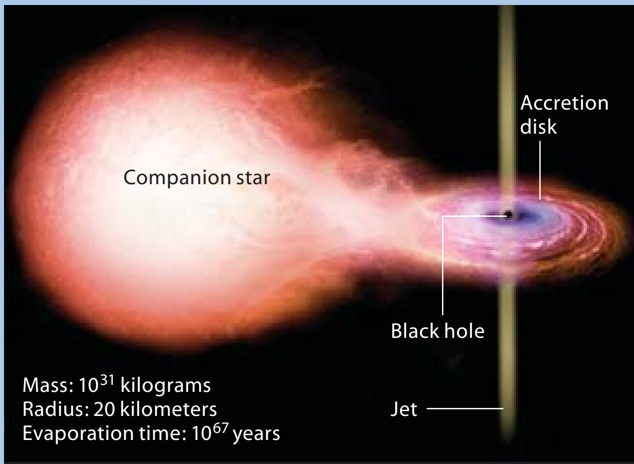
IN ITS MODERN FORM, the concept of black holes emerges from Einstein's general theory of relativity, which predicts that if matter is sufficiently compressed, its gravity becomes so strong that it carves out a region of space from which nothing can escape. The boundary of the region is the black hole's event horizon: objects can fall in, but none can come out. In the simplest case, where space has no hidden dimensions or those dimensions are smaller than the hole, its size is directly proportional to its mass. If you compressed the sun to a radius of three kilometers, about four millionths of its present size, it would become a black hole. For Earth to meet the same fate, you would need to squeeze it into a radius of nine millimeters, about a billionth its present size.

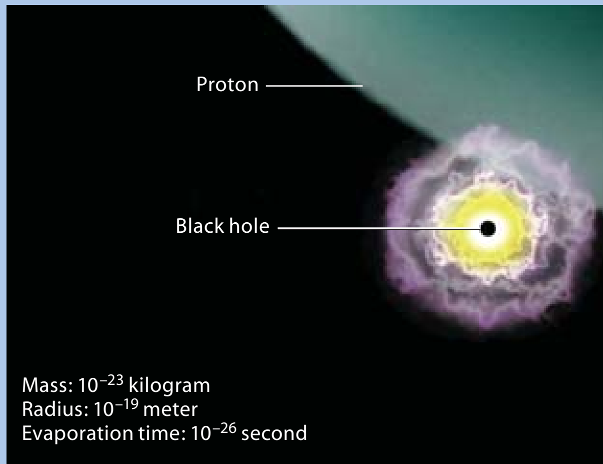
Thus, the smaller the hole, the higher the degree of compression that is required to create it. The density to which matter must be squeezed scales as the in-

JEAN-FRANCOIS PODEVIN



A Tale of Two Black Holes





ASTROPHYSICAL BLACK HOLES are thought to be the corpses of massive stars that collapsed under their own weight. As matter falls into them, they act like cosmic hydroelectric plants, releasing gravitational potential energy—the only power source that can account for the intense x-rays and gaseous jets that astronomers see spurting out of celestial systems such as the x-ray binary shown here.

MICROSCOPIC BLACK HOLES have masses ranging up to that of a large asteroid. They might have been churned out by the collapse of matter early in the big bang. If space has unseen extra dimensions, they might also be created by energetic particle collisions in today's universe. Rather than swallowing matter, they would give off radiation and decay away rapidly.

verse square of the mass. For a hole with the mass of the sun, the density is about 10^{19} kilograms per cubic meter, higher than that of an atomic nucleus. Such a density is about the highest that can be created through gravitational collapse in the present universe. A body lighter than the sun resists collapse because it gets stabilized by repulsive quantum forces between subatomic particles. Observationally, the lightest black hole candidates are about six solar masses.

Stellar collapse is not the only way that holes might form, however. In the early 1970s Stephen Hawking of the University of Cambridge and one of us (Carr) investigated a mechanism for generating holes in the early universe. These are termed "primordial" black holes. As the universe expands, the average density of matter decreases; therefore, the density was much higher in the past, in particular exceeding nuclear levels within the first microsecond of the big bang. The known laws of physics allow for a matter density up to the so-called Planck value of 10^{97} kilograms per cubic meter—the density at which the strength of gravity becomes so strong that quantum-mechanical fluctuations should break down the fabric of spacetime. Such a density would have been enough to create black holes a mere 10^{-35} meter across (a dimension known as the Planck length) with a mass of 10^{-8} kilogram (the Planck mass).

This is the lightest possible black hole according to conventional descriptions of gravity. It is much more massive but much smaller in size than an elementary particle. Progressively heavier primordial black holes could have formed as the cosmic density fell. Any lighter than 10^{12} kilograms would still be smaller than a proton, but beyond this mass the holes would be as large as more familiar physical objects. Those forming during the epoch when the cosmic density matched nuclear density would have a mass comparable to the sun's and so would be macroscopic.

The high densities of the early universe were a prerequisite for the formation of primordial black holes but did

Overview
Black Hole Factories

- Black holes need not be gargantuan, ravenous monsters. Theory implies that they can come in a huge variety of sizes, some even smaller than subatomic particles. Tiny holes should be wracked by quantum effects, and the very smallest would explode almost as soon as they formed.
- Small black holes might be left over from the early stages of the big bang, and astronomers might be able to detect some of them exploding today.
- Theorists have recently proposed that small black holes might be created in collisions in the present universe, even on Earth. They had thought that the requisite energies were too high, but if space has extra dimensions with the right properties, then the energy threshold for black hole production is much lower. If so, holes might be produced by the Large Hadron Collider (LHC) at CERN and in cosmic-ray collisions high in the atmosphere. Physicists could use the holes to probe the extra dimensions of space.

DON DIXON

not guarantee it. For a region to stop expanding and collapse to a black hole, it must have been denser than average, so density fluctuations were also necessary. Astronomers know that such fluctuations existed, at least on large scales, or else structures such as galaxies and clusters of galaxies would never have coalesced. For primordial black holes to form, these fluctuations must have been stronger on smaller scales than on large ones, which is possible though not inevitable. Even in the absence of fluctuations, holes might have formed spontaneously at various cosmological phase transitions—for example, when the universe ended its early period of accelerated expansion, known as inflation, or at the nuclear density epoch, when particles such as protons condensed out of the soup of their constituent quarks. Indeed, cosmologists can place important constraints on models of the early universe from the fact that not too much matter ended up in primordial black holes.

Going, Going, Gone?

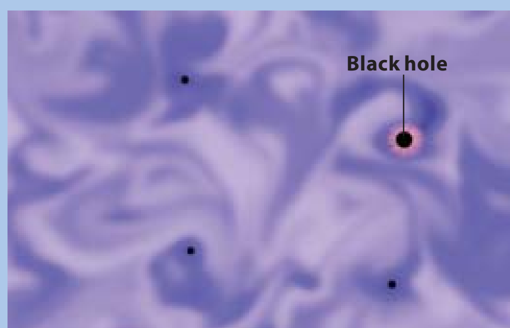
THE REALIZATION that holes could be small prompted Hawking to consider what quantum effects might come into play, and in 1974 he came to his famous conclusion that black holes do not just swallow particles but also spit them out. Hawking predicted that a hole radiates thermally like a hot coal, with a temperature inversely proportional to its mass. For a solar-mass hole, the temperature is around a millionth of a kelvin, which is completely negligible in today's universe. But for a black hole of 10^{12} kilograms, which is about the mass of a mountain, it is 10^{12} kelvins—hot enough to emit both massless particles, such as photons, and massive ones, such as electrons and positrons.

Because the emission carries off energy, the mass of the hole tends to decrease. So a black hole is highly unstable. As it shrinks, it gets steadily hotter, emitting increasingly energetic particles and shrinking faster and faster. When the hole shrivels to a mass of about 10^6 kilograms, the game is up: within a second, it explodes with the energy of a million-megaton nuclear bomb. The total time for a black hole to evaporate away is proportional to the cube of its initial mass. For a solar-mass hole, the lifetime is an unobservably long 10^{64} years. For a 10^{12}-kilogram one, it is 10^{10} years—about the present age of the universe. Hence, any primordial black holes of this mass would be completing their evaporation and exploding right now. Any smaller ones would have evaporated during an earlier cosmological epoch.

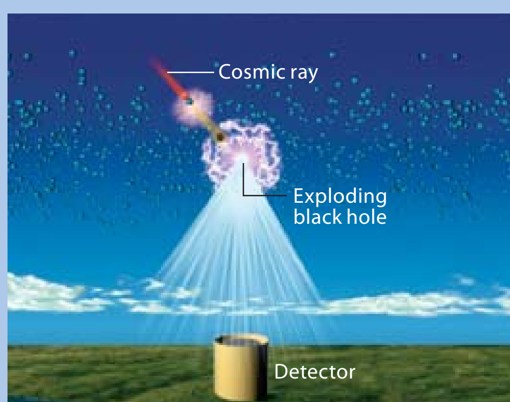
Hawking's work was a tremendous conceptual advance because it linked three previously disparate areas of physics: general relativity, quantum theory and thermodynamics. It was also a step toward a full quantum theory of gravity. Even if primordial black holes never actually formed, thinking about them has led to remarkable physical insights. So it can be useful to study something even if it does not exist.

In particular, the discovery opened up a profound paradox that aims at the heart of why general relativity and quantum mechanics are so hard to reconcile. According to relativity theory, information about what falls into a black hole is forever lost. If the hole evaporates, however, what happens to the information contained within? Hawking suggested that black holes completely evaporate, destroying the information and violating the basic tenets of quantum mechanics. But such destruction of information also conflicts with the law of

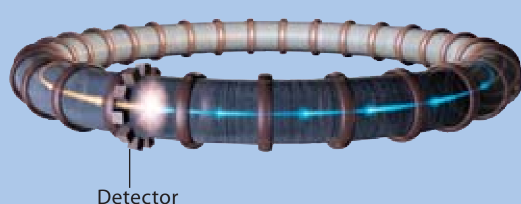
Ways to Make a Mini Black Hole



Primordial Density Fluctuations
Early in the history of our universe, space was filled with hot, dense plasma. The density varied from place to place, and in locations where the relative density was sufficiently high, the plasma could collapse into a black hole.



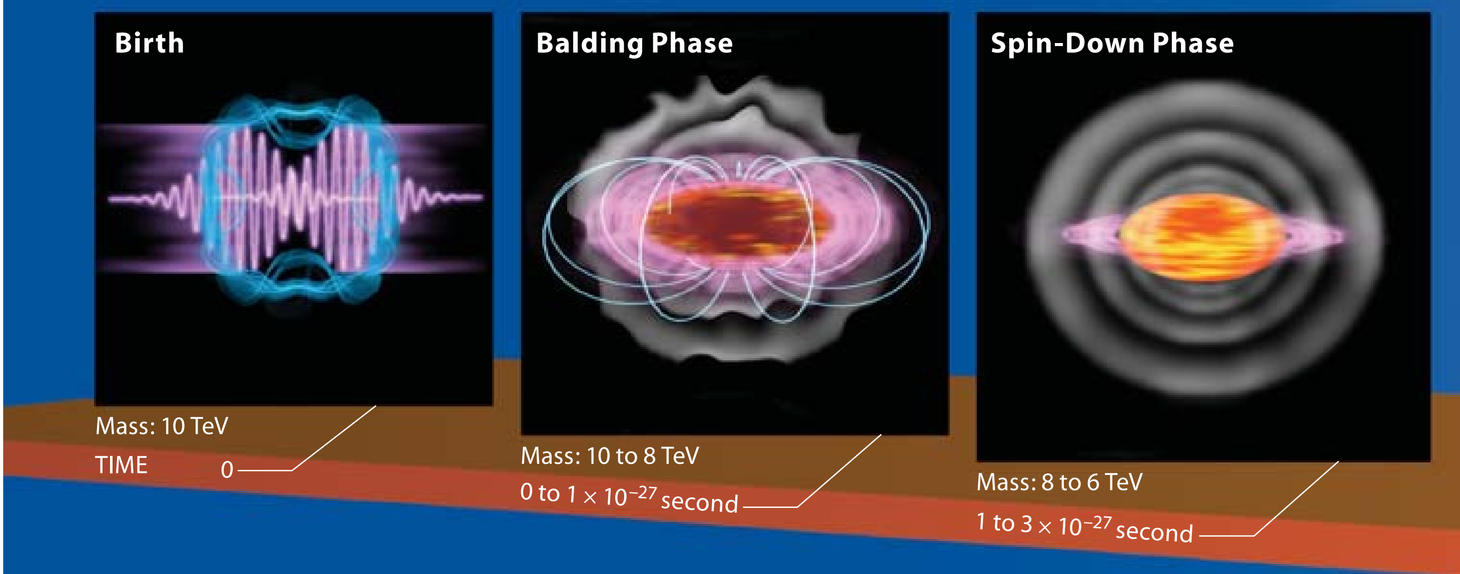
Cosmic-Ray Collisions
Cosmic rays—highly energetic particles from celestial sources—could smack into Earth's atmosphere and form black holes. They would explode in a shower of radiation and secondary particles that could be detected on the ground.



Particle Accelerator
An accelerator such as the LHC could crash two particles together at such an energy that they would collapse into a black hole. Detectors would register the subsequent decay of the hole.

DON DIXON

The Rise and Demise of a Quantum Black Hole



If conditions are right, two particles (*shown here as wave packets*) can collide to create a black hole. The newborn hole is asymmetrical. It can be rotating, vibrating and electrically charged. (Times and masses are approximate; 1 TeV is the energy equivalent of about 10^{-24} kilogram.)

As it settles down, the black hole emits gravitational and electromagnetic waves. To paraphrase physicist John A. Wheeler, the hole loses its hair—it becomes an almost featureless body, characterized solely by charge, spin and mass. Even the charge quickly leaks away as the hole gives off charged particles.

The black hole is no longer black: it radiates. At first, the emission comes at the expense of spin, so the hole slows down and relaxes into a spherical shape. The radiation emerges mainly along the equatorial plane of the black hole.

DON DIXON; C. M. HARRIS *University of Cambridge* AND THE ATLAS COLLABORATION (*far right*)

energy conservation, making this possibility implausible.

One alternative, that evaporating black holes leave behind remnants, is equally unpalatable. For these remnants to encode all the information that could have gone into the black hole, they would have to come in an infinite variety of types. The laws of physics predict that the rate of production of a particle is proportional to the number of types of that particle. Therefore, the black hole remnants would be produced at an infinite rate; even such everyday physical processes as turning on a microwave oven would generate them. Nature would be catastrophically unstable. A third (and most likely) possibility is that information escapes through a breakdown of locality—the notion that events at spatially separated points can influence one another only after light has had time to travel between them—that is more profound than ordinary quantum nonlocality. This conundrum challenges theorists to this day.

Looking for Holes

PROGRESS IN PHYSICS usually requires some guidance from experiment, so the questions raised by microscopic black holes motivate an empirical search for them. One possibility is that astronomers might be able to detect primordial black holes with an initial mass of 10^{12} kilograms exploding in the present universe. Roughly a tenth of the mass of these holes would go into gamma rays. In 1976 Hawking and Don Page, then at the California Institute of Technology, realized that gamma-ray background observations place stringent upper limits on the number of such holes. They could not, for example, constitute a significant amount of the universe's dark matter, and their explosions would rarely be close enough to be detectable. In the mid-1990s, however, David Cline of the University of California, Los Angeles, and his colleagues suggested that the shortest gamma-ray bursts might be primordial black holes blowing up. Although longer bursts are thought to be associated with exploding or merging stars, the short events may have another explanation. Future observations should settle this issue, but the possibility that astronomical observations could probe the final stages of black hole evaporation is tantalizing.

The production of black holes by particle accelerators is an even more exciting possibility. When it comes to producing high densities, no device outdoes accelerators such as the LHC and the Tevatron at Fermi National Accelerator Laboratory in Batavia, Ill. These ma-

Mass: 6 to 2 TeV

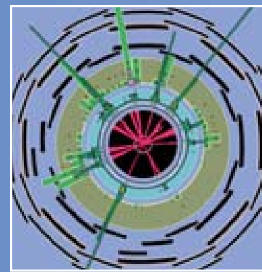
3 to 20 × 10^{-27} second

Having lost its spin, the black hole is now an even simpler body than before, characterized solely by mass. Even the mass leaks away in the form of radiation and massive particles, which emerge in every direction.

Mass: 2 to 0 TeV

20 to 22 × 10^{-27} second

The hole approaches the Planck mass—the lowest mass possible for a hole, according to present theory—and winks into nothingness. String theory suggests that the hole would begin to emit strings, the most fundamental units of matter.



SIMULATED DECAY of a black hole shows a particle accelerator and detector in cross section. From the center of the accelerator pipe (*black circle*) emerge particles (*spokes*) registered by layers of detectors (*concentric colored rings*).

chines accelerate subatomic particles, such as protons, to velocities exceedingly close to the speed of light. The particles then have enormous kinetic energies. At the LHC, a proton will reach an energy of roughly seven tera-electron volts (TeV). In accord with Einstein's famous relation $E = mc^2$, this energy is equivalent to a mass of 10^{-23} kilogram, or 7,000 times the proton's rest mass. When two such particles collide at close range, their energy is concentrated into a tiny region of space. So one might guess that, once in a while, the colliding particles will get close enough to form a black hole.

As it stands, this argument has a problem: a mass of 10^{-23} kilogram is far shy of the Planck value of 10^{-8} kilogram, which conventional gravity theory implies is the lightest possible hole. This lower limit arises from the uncertainty principle of quantum mechanics. Because particles also behave like waves, they are smeared out over a distance that decreases with increasing energy—at LHC energies, about 10^{-19} meter. So this is the smallest region into which a particle's energy can be packed. It allows for a density of 10^{34} kilograms per cubic meter, which is high but not high enough to create a hole. For a particle to be both energetic enough and compact enough to form a black hole, it must have the Planck energy, a factor of 10^{15} beyond the energy of the LHC.

Interestingly, accelerators may be able to create objects mathematically related to black holes. The Relativistic Heavy Ion Collider at Brookhaven National Laboratory in Upton, N.Y., may already have done so, but black holes themselves appear to lie out of reach.

Reaching into Other Dimensions

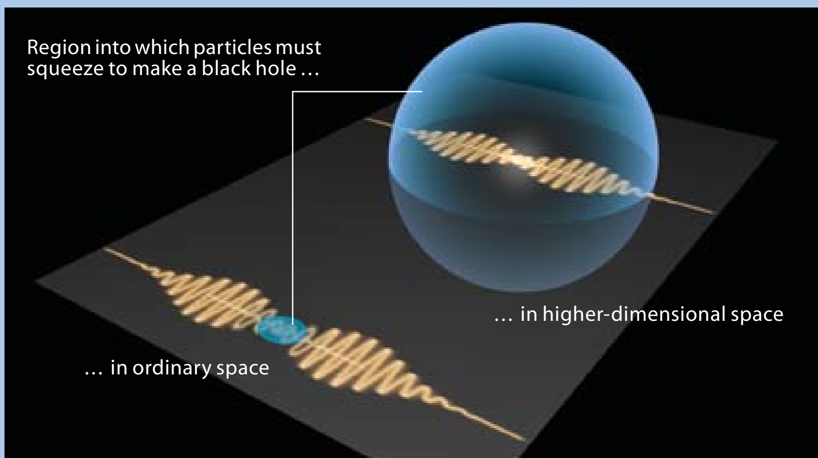
OVER THE PAST DECADE, however, physicists have realized that the stan-

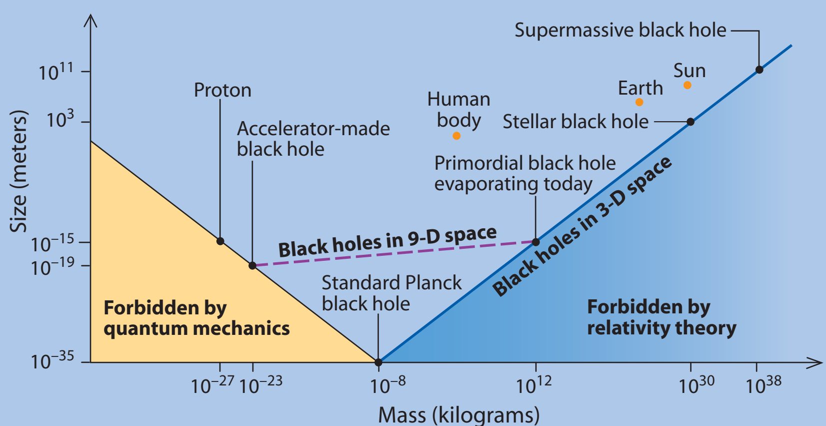
THE AUTHORS

BERNARD J. CARR and *STEVEN B. GIDDINGS* first met in person at a conference to celebrate Stephen Hawking's 60th birthday in 2002. Carr traces his enthusiasm for astrophysics to the famous 1969 BBC television documentary by Nigel Calder entitled *The Violent Universe*. He became a graduate student of Hawking's in the 1970s, was one of the first scientists to investigate small black holes and today is professor at Queen Mary, University of London. Giddings caught the physics bug when his father first told him about the weird properties of quantum mechanics. He went on to become an expert on quantum gravity and cosmology, was among the first to study the possibility of creating black holes in particle accelerators and is now professor at the University of California, Santa Barbara. When not theorizing about gravity, he struggles against it by rock climbing.

Making Holes Is Hard to Do

How much do you need to squeeze a piece of matter to turn it into a black hole? The lighter a body is, the more you must compress it before its gravity becomes strong enough to make a hole. Planets and people are farther from the brink than stars are (*graph*). The wave nature of matter resists compression; particles cannot be squeezed into a region smaller than their characteristic wavelength (*diagram*), suggesting that no hole could be smaller than 10^{-8} kilogram. But if space has extra dimensions, gravity would be inherently stronger over short distances and an object would not need to be squeezed as much, giving would-be hole makers hope that they might succeed in the near future.





dard estimate of the necessary Planckian density could be too high. String theory, one of the leading contenders for a quantum theory of gravity, predicts that space has dimensions beyond the usual three. Gravity, unlike other forces, should propagate into these dimensions and, as a result, grow unexpectedly stronger at short distances. In three dimensions, the force of gravity quadruples as you halve the distance between two objects. But in nine dimensions, gravity would get 256 times as strong. This effect can be quite important if the extra dimensions of space are sufficiently large, and it has been widely investigated in the past few years. There are also other configurations of extra dimensions, known as warped compactifications, that have the same gravity-magnifying effect and may be even more likely to occur if string theory is correct; these have been extensively studied in recent years.

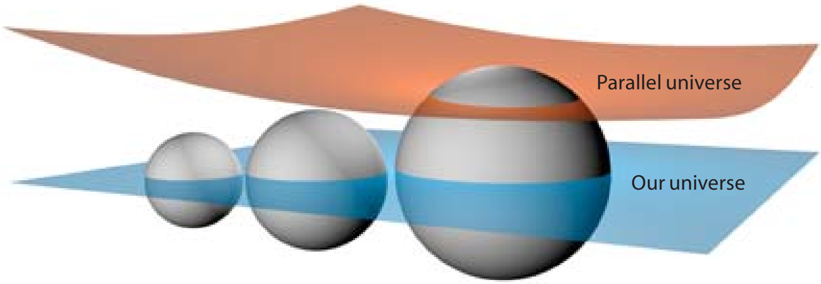
This enhanced growth of the strength of gravity means that the true energy scale at which the laws of gravity and quantum mechanics clash—and black holes can be made—could be much lower than the traditional expectation. Although no experimental evidence yet supports this possibility, the idea sheds new light on various theoretical conundrums. And if it is true, the density required to create black holes could lie within the range of the LHC.

The theoretical study of black hole production in high-energy collisions goes back to the work of Roger Penrose of the University of Oxford in the mid-1970s and Peter D'Eath and Philip Norbert Payne, both then at Cambridge, in the early 1990s. The newfound possibility of large extra dimensions breathed new life into these investigations and motivated Tom Banks of the University of California, Santa Cruz, and Rutgers University and Willy Fischler of the University of Texas at Austin to write a 1999 paper with a preliminary discussion of black hole production.

At a 2001 workshop two groups—one of us (Giddings), with Scott Thomas, then at Stanford University, and Savas Dimopoulos of Stanford, with Greg Landsberg of Brown University—independently described the observable effects, and thus the potential for discovery, of black hole production at particle colliders such as the LHC. After a few calculations, we were astounded. Rough estimates indicated that under the most optimistic scenarios, corresponding to the lowest plausible value for the Planck scale, black holes could be produced at the rate of one per second. Physicists refer to an accelerator producing a particle at this rate as a "factory," so the LHC would be a black hole factory.

The evaporation of these holes would leave very distinctive imprints on the detectors. Typical collisions produce moderate numbers of high-energy particles, but a decaying black hole is different. According to Hawking's work, it radiates a large number of particles in all directions with very high energies. The decay products include all the particle species found in nature. Several research groups have since done increasingly detailed investigations into the remarkable signatures that black holes would produce in the detectors at the LHC.

DON DIXON; JANA BRENNING (*graph*)



BLACK HOLES OF DIFFERENT SIZES could probe extra dimensions that are otherwise inaccessible to us. Because gravity, unlike other forces, extends into those dimensions, so do black holes. Physicists would vary their size by tuning the particle accelerator to different energies. If a hole intersects a parallel universe, it will decay faster and appear to give off less energy (because some of the energy is absorbed by that other universe).

DON DIXON

Is It Raining Black Holes?

THE PROSPECT of producing black holes on Earth may strike some as folly. How do we know that they would safely decay, as Hawking predicted, instead of continuing to grow, eventually consuming the entire planet? At first glance, this seems like a serious concern, especially given that some details of Hawking's original argument may be incorrect—specifically the claim that information is destroyed in black holes.

But general quantum reasoning implies that microscopic black holes cannot be stable and therefore are safe. Concentrations of mass energy, such as elementary particles, are stable only if a conservation law forbids their decay; examples include the conservation of electric charge and of baryon number (which, unless it is somehow violated, assures the stability of protons). There is no such conservation law to stabilize a small black hole. In quantum theory, anything not expressly forbidden is compulsory, so small black holes will rapidly decay, in accord with the second law of thermodynamics.

Indeed, high-energy collisions such as those at the LHC have already taken place—for example, in the early universe and even now, when sufficiently high energy cosmic rays hit our atmosphere. So if collisions at LHC energies can make black holes, nature has already been harmlessly producing them right over our heads. Early estimates by Giddings and Thomas indicated that the highest-energy cosmic rays—protons or heavier atomic nuclei with energies of up to 10^9 TeV—could produce as many as 100 black holes in the atmosphere a year.

In addition, they—along with David Dorfan of U.C. Santa Cruz and Tom Rizzo of the Stanford Linear Accelerator Center and, independently, Jonathan L. Feng of the University of California, Irvine, and Alfred D. Shapere of the University of Kentucky—have discovered that collisions of cosmic neutrinos might be even more productive. If so, the new Auger cosmic-ray observatory in Argentina, which is now taking data, and the upgraded Fly's Eye observatory in Utah may be able to see upward of several holes a year. These observations, however, would not obviate the need for accelerator experiments, which could generate holes more reliably, in greater numbers and under more controlled circumstances.

Producing black holes would open up a whole new frontier of physics. Their mere presence would be compelling evidence of the previously hidden dimensions of space, and by observing their properties, physicists might begin to explore the geographic features of those dimensions. For example, as accelerators manufacture black holes of increasing mass, the holes would poke further into the extra dimensions and could become comparable in size to one or more of them, leading to a distinctive change in the dependence of a hole's temperature on mass. Likewise, if a black hole grows large enough to intersect a parallel three-dimensional universe in the extra dimensions, its decay properties would suddenly change.

Producing black holes in accelerators would also represent the end of one of humankind's historic quests: understanding matter on ever finer scales. Over the past century, physicists have pushed back the frontier of the small—from dust motes to atoms to protons and neutrons to quarks. If they can create black holes, they will have reached the Planck scale, which is believed to be the shortest meaningful length, the limiting distance below which the very notions of space and length probably cease to exist. Any attempt to investigate the possible existence of shorter distances, by performing higher-energy collisions, would inevitably result in black hole production. Higher-energy collisions, rather than splitting matter into finer pieces, would simply produce bigger black holes. In this way, the appearance of black holes would mark the close of a frontier of science. In its place, however, would be a new frontier, that of exploring the geography of the extra dimensions of space. SA

MORE TO EXPLORE

Black Holes and Time Warps: Einstein's Outrageous Legacy. Kip S. Thorne. W. W. Norton, 1995.

High Energy Colliders as Black Hole Factories: The End of Short Distance Physics. Steven B. Giddings and Scott Thomas in *Physical Review D*, Vol. 65, Paper No. 056010; 2002. **arxiv.org/abs/hep-ph/0106219**

Black Holes at the LHC. Savas Dimopoulos and Greg Landsberg in *Physical Review Letters*, Vol. 87, Paper No. 161602; 2001. **hep-ph/0106295**

Black Holes from Cosmic Rays: Probes of Extra Dimensions and New Limits on TeV-Scale Gravity. Luis A. Anchordoqui, Jonathan L. Feng, Haim Goldberg and Alfred D. Shapere in *Physical Review D*, Vol. 65, Paper No. 124027; 2002. **hep-ph/0112247**

Black Holes at Accelerators. Steven B. Giddings in *The Future of Theoretical Physics and Cosmology*. Edited by G. W. Gibbons, E.P.S. Shellard and S. J. Rankin. Cambridge University Press, 2003. **hep-th/0205027**

Primordial Black Holes. Bernard Carr. Ibid. Similar paper available at **astro-ph/0310838**

how to build a TIME MACHINE

It wouldn't be easy, but it might be possible

By Paul Davies

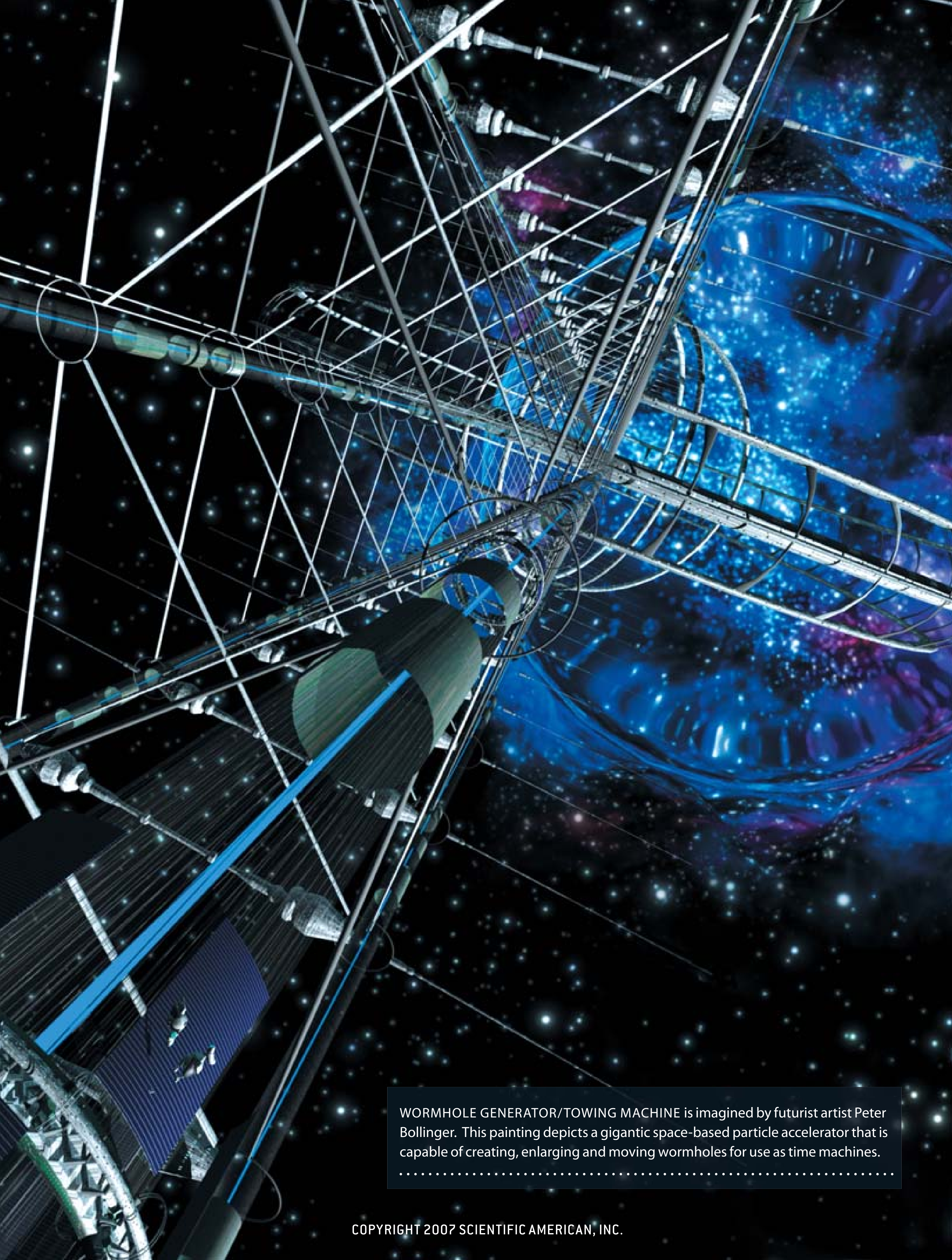
Time travel has been a popular science-fiction theme since H. G. Wells wrote his celebrated novel *The Time Machine* in 1895. But can it really be done? Is it possible to build a machine that would transport a human being into the past or future?

For decades, time travel lay beyond the fringe of respectable science. In recent years, however, the topic has become something of a cottage industry among theoretical physicists. The motivation has been partly recreational—time travel is fun to think about. But this research has a serious side, too. Understanding the relation between cause and effect is a key part of attempts to construct a unified theory of physics. If unrestricted time travel were possible, even in principle, the nature of such a unified theory could be drastically affected.

Our best understanding of time comes from Einstein's theories of relativity. Prior to these theories, time was widely regarded as absolute and universal, the same for everyone no matter what their physical circumstances were. In his special theory of relativity, Einstein proposed that the measured interval between two events depends on how the observer is moving. Crucially, two observers who move differently will experience different durations between the same two events.

The effect is often described using the "twin paradox." Suppose that Sally and Sam are twins. Sally boards a rocket ship and travels at high speed to a nearby star, turns around and flies back to Earth, while Sam stays at home. For Sally the duration of the journey might be, say, one year, but when she returns and steps out of the spaceship, she finds that 10 years have elapsed on Earth. Her brother is now nine years older than she is. Sally and Sam are no longer the same age, despite the fact that they were born on the same day. This example illustrates a limited type of time travel. In effect, Sally has leaped nine years into Earth's future.


PETER BOLLINGER



WORMHOLE GENERATOR/TOWING MACHINE is imagined by futurist artist Peter Bollinger. This painting depicts a gigantic space-based particle accelerator that is capable of creating, enlarging and moving wormholes for use as time machines.

The **wormhole** was used as a fictional device by Carl Sagan in his novel *Contact*.

Jet Lag

THE EFFECT, KNOWN AS time dilation, occurs whenever two observers move relative to each other. In daily life we don't notice weird time warps, because the effect becomes dramatic only when the motion occurs at close to the speed of light. Even at aircraft speeds, the time dilation in a typical journey amounts to just a few nanoseconds—hardly an adventure of Wellsian proportions. Nevertheless, atomic clocks are accurate enough to record the shift and confirm that time really is stretched by motion. So travel into the future is a proved fact, even if it has so far been in rather unexciting amounts.

To observe really dramatic time warps, one has to look beyond the realm of ordinary experience. Subatomic particles can be propelled at nearly the speed of light in large accelerator machines. Some of these particles, such as muons, have a built-in clock because they decay with a definite half-life; in accordance with Einstein's theory, fast-moving muons inside accelerators are observed to decay in slow motion. Some cosmic rays also experience spectacular time warps. These particles move so close to the speed of light that, from their point of view, they cross the galaxy in minutes, even though in Earth's frame of reference they seem to take tens of thousands of years. If time dilation did not occur, those particles would never make it here.

Speed is one way to jump ahead in time. Gravity is another. In his general theory of relativity, Einstein predicted that gravity slows time. Clocks run a bit faster in the attic than in the basement, which is closer to the center of Earth and therefore deeper down in a gravitational field. Similarly, clocks run faster in space than on the ground. Once again the effect is minuscule, but it has been directly measured using accurate clocks. Indeed, these time-warping effects have to be taken into account in the Global Positioning System. If they weren't, sailors, taxi drivers and cruise missiles could find themselves many kilometers off course.

At the surface of a neutron star, gravity is so strong that time is slowed by about 30 percent relative to Earth time. Viewed from such a star, events here would resemble a fast-forwarded video. A black hole represents the ultimate time warp; at the surface of the hole, time stands still relative to Earth. This means that if you fell into a black hole from nearby, in the brief interval it took you to reach the surface, all of eternity would pass by in the wider universe. The region within the black hole is therefore beyond the end of time, as far as the outside universe is concerned. If an astronaut could zoom very close to a black hole and return unscathed—admittedly a fanciful, not to mention foolhardy, prospect—he could leap far into the future.

My Head Is Spinning

SO FAR I HAVE DISCUSSED travel forward in time. What about going backward? This is much more problematic. In 1948 Kurt Gödel of the Institute for Advanced Study in Princeton, N.J., produced a solution of Einstein's gravitational field equations that described a rotating universe. In this universe, an astronaut could travel through space so as to reach his own past. This comes about because of the way gravity affects light. The rotation of the universe would drag light (and thus the causal relations between objects) around with it, enabling a material object to travel in a closed loop in space that is also a closed loop in time, without at any stage exceeding the speed of light in the immediate neighborhood of the particle. Gödel's solution was shrugged aside as a mathematical curiosity—after all, observations show no sign that the universe as a whole is spinning. His result served nonetheless to demonstrate that going back in time was not forbidden by the theory of relativity. Indeed, Einstein confessed that he was troubled by the thought that his theory might permit travel into the past under some circumstances.

Other scenarios have been found to permit travel into the past. For example, in 1974 Frank J. Tipler of Tulane University calculated that a massive, infinitely long cylinder spinning on its axis at near the speed of light could let astronauts visit their own past, again by dragging light around the cylinder into a loop. In 1991 J. Richard Gott of Princeton University predicted that cosmic strings—structures that cosmologists think were created in the early stages of the big

EVERETT COLLECTION

Overview
Time Travel

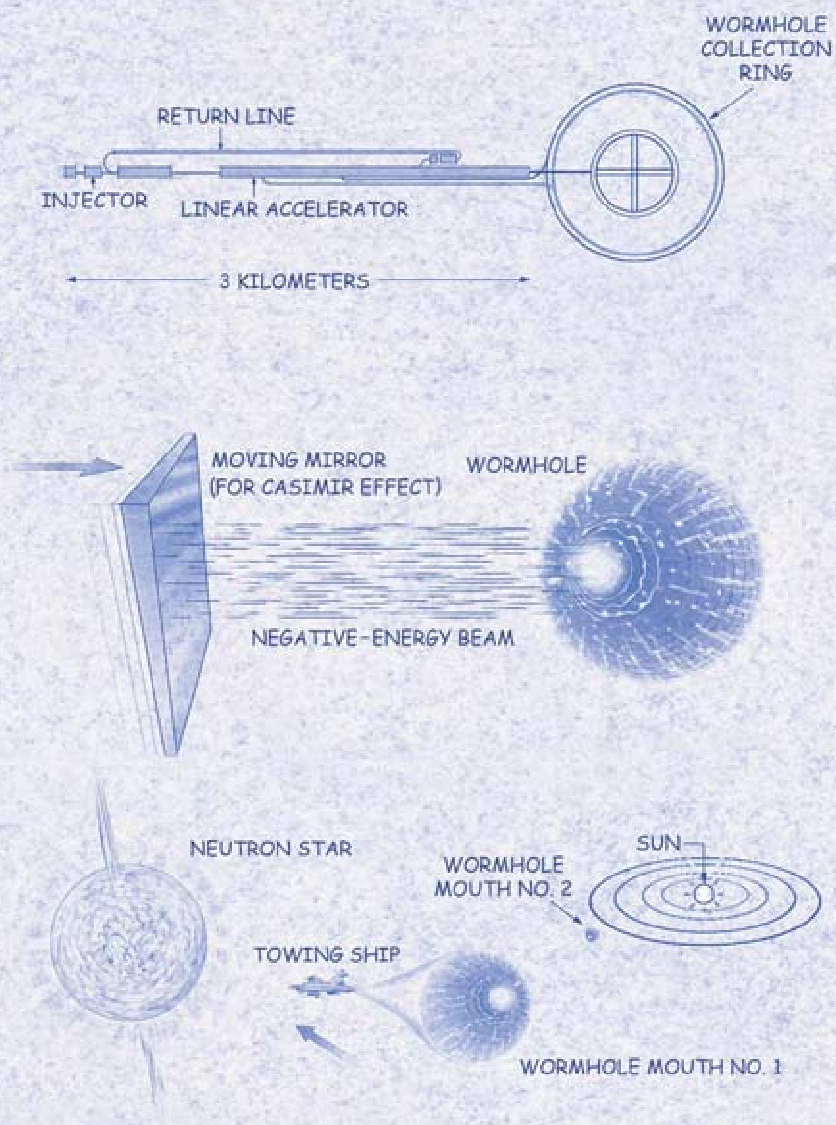
- Traveling forward in time is easy enough. If you move close to the speed of light or sit in a strong gravitational field, you experience time more slowly than other people do—another way of saying that you travel into their future.
- Traveling into the past is rather trickier. Relativity theory allows it in certain spacetime configurations: a rotating universe, a rotating cylinder and, most famously, a wormhole—a tunnel through space and time.

A Wormhole Time Machine in Three Not So Easy Steps

1 Find or build a wormhole—a tunnel connecting two different locations in space. Large wormholes might exist naturally in deep space, a relic of the big bang. Otherwise we would have to make do with subatomic wormholes, either natural ones (which are thought to be winking in and out of existence all around us) or artificial ones (produced by particle accelerators, as imagined here). These smaller wormholes would have to be enlarged to a useful size, perhaps using energy fields like those that caused space to inflate shortly after the big bang.

2 Stabilize the wormhole. An infusion of negative energy, produced by quantum means such as the so-called Casimir effect, would allow a signal or object to pass safely through the wormhole. Negative energy counteracts the tendency of the wormhole to pinch off into a point of infinite or near-infinite density. In other words, it prevents the wormhole from becoming a black hole.

3 Tow the wormhole. A spaceship, presumably of highly advanced technology, would separate the mouths of the wormhole. One mouth might be positioned near the surface of a neutron star, an extremely dense star with a strong gravitational field. The intense gravity causes time to pass more slowly. Because time passes more quickly at the other wormhole mouth, the two mouths become separated not only in space but also in time.



PHILIP HOWE

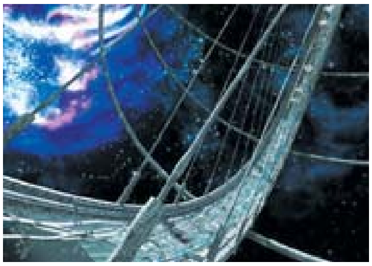
bang—could produce similar results. But in the mid-1980s the most realistic scenario for a time machine emerged, based on the concept of a wormhole.

In science fiction, wormholes are sometimes called stargates; they offer a shortcut between two widely separated points in space. Jump through a hypothetical wormhole, and you might come out moments later on the other side of the galaxy. Wormholes naturally fit into the general theory of relativity, whereby gravity warps not only time but also space. The theory allows the analogue of alternative road and tunnel routes connecting two points in space. Mathematicians refer to such a space as multiply connected. Just as a tunnel passing under a hill can be shorter than the surface street, a wormhole may be shorter than the usual route through ordinary space.

The wormhole was used as a fictional device by Carl Sagan in his 1985 novel *Contact*. Prompted by Sagan, Kip S. Thorne and his co-workers at the California Institute of

THE AUTHOR

PAUL DAVIES is director of Beyond: Center for Fundamental Concepts in Science at Arizona State University. A theoretical physicist and cosmologist by profession, he also works in the field of astrobiology. He is one of the most prolific writers of popular-level books in physics. His scientific research interests include black holes, quantum field theory, the nature of consciousness, and the origin and evolution of life.



Unlike a black hole, which offers a **one-way journey to nowhere,** a wormhole would have an exit as well as an entrance.

Technology set out to find whether wormholes were consistent with known physics. Their starting point was that a wormhole would resemble a black hole in being an object with fearsome gravity. But unlike a black hole, which offers a one-way journey to nowhere, a wormhole would have an exit as well as an entrance.

In the Loop

FOR THE WORMHOLE to be traversable, it must contain what Thorne termed exotic matter. In effect, this is something that will generate antigravity to combat the natural tendency of a massive system to implode into a black hole under its intense weight. Antigravity, or gravitational repulsion, can be generated by negative energy or pressure. Negative-energy states are known to exist in certain quantum systems, which suggests that Thorne's exotic matter is not ruled out by the laws of physics, although it is unclear whether enough antigravitating stuff can be assembled to stabilize a wormhole.

Soon Thorne and his colleagues realized that if a stable wormhole could be created, then it could readily be turned into a time machine. An astronaut who passed through one might come out not only somewhere else in the universe but somewhen else, too—in either the future or the past.

To adapt the wormhole for time travel, one of its mouths could be towed to a neutron star and placed close to its surface. The gravity of the star would slow time near that wormhole mouth, so that a time difference between the ends of the wormhole would gradually accumulate. If both mouths were then parked at a convenient place in space, this time difference would remain frozen in.

Suppose the difference were 10 years. An astronaut passing through the wormhole in one direction would jump 10 years into the future, whereas an astronaut passing in the other direction would jump 10 years into the past. By returning to his starting point at high speed across ordinary space, the second astronaut might get back home before he left. In other words, a closed loop in space could become a loop in time as well. The one restriction is that the astronaut could not return to a time before the wormhole was first built.

A formidable problem that stands in the way of making a wormhole time machine is the creation of the wormhole in the first place. Possibly space is threaded with such structures naturally—relics of the big bang. If so, a supercivilization might commandeer one. Alternatively, wormholes might naturally come into existence on tiny scales, the so-called Planck length, about 20 factors of 10 as small as an atomic nucleus. In principle, such a minute wormhole could be stabilized by a pulse of energy and then somehow inflated to usable dimensions.

Existing Forms of Forward Travel

System	Specifications	Cumulative Time Lag
Airline flight	920 kilometers an hour for eight hours	10 nanoseconds (relative to inertial reference frame)
Nuclear submarine tour	300 meters' depth for six months	500 nanoseconds (relative to sea level)
Cosmic-ray neutron	10^{18} electron volts	Mean life stretched from 15 minutes to 30,000 years
Neutron star	Redshift 0.2	Time intervals expand 30 percent (relative to deep space)

Censored!

ASSUMING THAT the engineering problems could be overcome, the production of a time machine could open up a Pandora's box of causal paradoxes. Consider, for example, the time traveler who visits the past and murders his mother when she was a young girl. How do we make sense of this? If the girl dies, she cannot become the time traveler's mother. But if the time traveler was never born, he could not go back and murder his mother.

Paradoxes of this kind arise when the time traveler tries to change the past, which is obviously impossible. But that does not prevent someone from being a part of the past. Suppose the time traveler goes back and rescues a young girl from murder, and this girl grows up to become his mother. The causal loop is now self-consistent and no longer paradoxical. Causal consistency might impose restrictions on what a time traveler is able to do, but it does not rule out time travel per se.

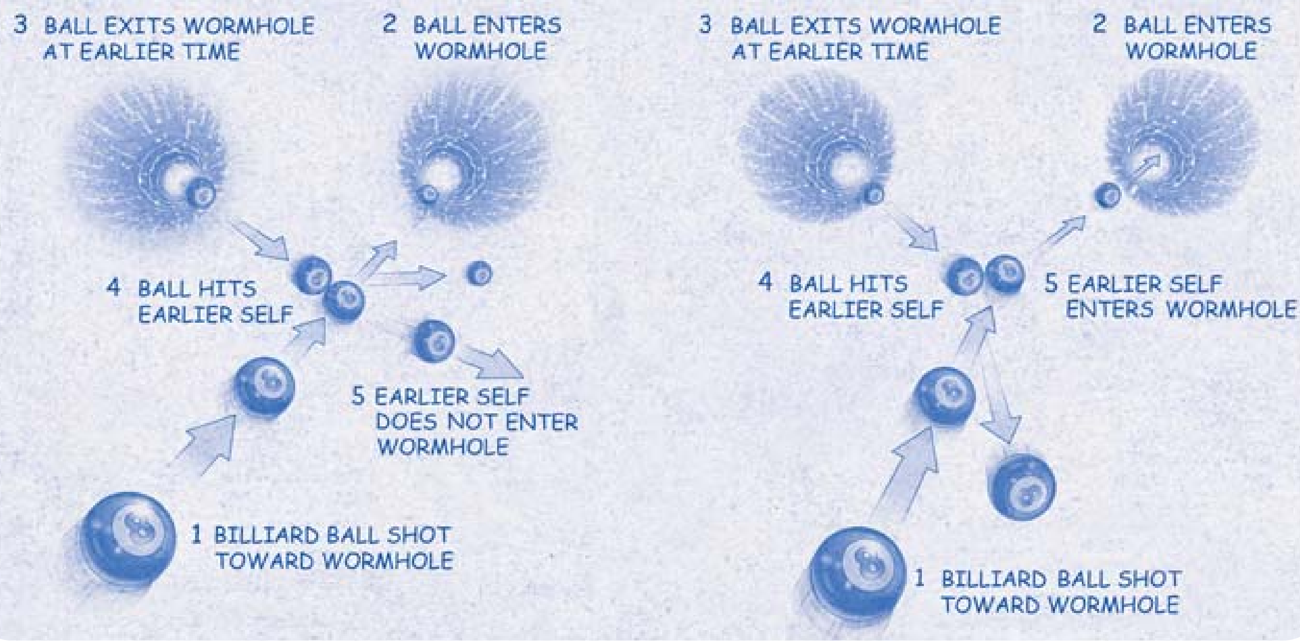
Even if time travel isn't strictly paradoxical, it is certainly weird. Consider the time traveler who leaps ahead a year and

PETER BOLLINGER

Mother of All Paradoxes

The notorious mother paradox (sometimes formulated using other familial relationships) arises when people or objects can travel backward in time and alter the past. A simplified version involves billiard balls. A billiard ball passes through a wormhole time machine. When it emerges, it hits its earlier self, thereby preventing it from ever entering the wormhole.

Resolution of the paradox proceeds from a simple realization: the billiard ball cannot do something that is inconsistent with logic or with the laws of physics. It cannot pass through the wormhole in such a way that will prevent it from passing through the wormhole. But nothing stops it from passing through the wormhole in an infinity of other ways.



PHILIP HOWE

reads about a new mathematical theorem in a future edition of *Scientific American*. He notes the details, returns to his own time and teaches the theorem to a student, who then writes it up for *Scientific American*. The article is, of course, the very one that the time traveler read. The question then arises: Where did the information about the theorem come from? Not from the time traveler, because he read it, but not from the student either, who learned it from the time traveler. The information seemingly came into existence from nowhere, reasonlessly.

The bizarre consequences of time travel have led some scientists to reject the notion outright. Stephen Hawking of the University of Cambridge has proposed a "chronology protection conjecture," which would outlaw causal loops. Because the theory of relativity is known to permit causal loops, chronology protection would require some other factor to intercede to prevent travel into the past. What might this factor be? One suggestion is that quantum processes will come to the rescue. The existence of a time machine would allow particles to loop into their own past. Calculations hint that the ensuing disturbance would become self-reinforcing, creating a runaway surge of energy that would wreck the wormhole.

Chronology protection is still just a conjecture, so time travel remains a possibility. A final resolution of the matter may have to await the successful union of quantum mechanics and gravitation, perhaps through a theory such as string theory or its extension, so-called M-theory. It is even conceivable that the next generation of particle accelerators will be able to create subatomic wormholes that survive long enough for nearby particles to execute fleeting causal loops. This would be a far cry from Wells's vision of a time machine, but it would forever change our picture of physical reality. SA

MORE TO EXPLORE

Time Machines: Time Travel in Physics, Metaphysics, and Science Fiction. Paul J. Nahin. American Institute of Physics, 1993.

The Quantum Physics of Time Travel. David Deutsch and Michael Lockwood in *Scientific American*, Vol. 270, No. 3, pages 68–74; March 1994.

Black Holes and Time Warps: Einstein's Outrageous Legacy. Kip S. Thorne. W. W. Norton, 1994.

Time Travel in Einstein's Universe: The Physical Possibilities of Travel through Time. J. Richard Gott III. Houghton Mifflin, 2001.

How to Build a Time Machine. Paul Davies. Viking, 2002.

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

THE BRIGHTEST EXPLOSIONS in the universe

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

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

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

The flash was another of the famous gamma-ray bursts, which in recent decades have been one of astronomy's most intriguing mysteries. The first sighting of a gamma-ray burst (GRB) came on July 2, 1967, from military satellites watching for nuclear tests in space. These cosmic explosions proved to be rather different from the man-made explosions that the satellites were designed to detect. For most of the next 30 years, each new burst had merely heightened the puzzlement. Whenever researchers thought they had the explanation, the evidence sent them back to square one.

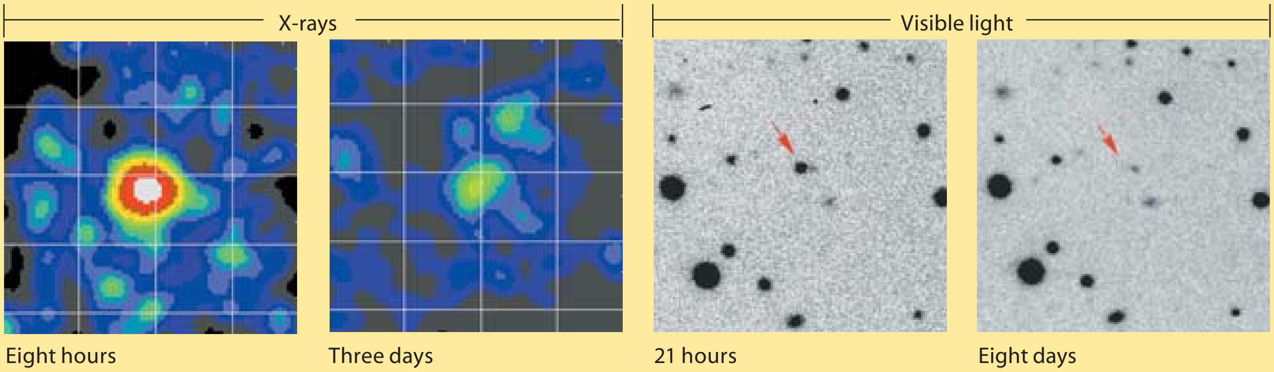
The monumental discoveries of the past decade have brought astronomers closer to a definitive answer. Before 1997, most of what we knew about GRBs was based on observations from the Burst and Transient Source Experiment (BATSE) onboard the Compton Gamma Ray Observatory. BATSE revealed that two or three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

Arguably the most important result from BATSE concerned the distribution of the bursts. They occur isotropically—that is, they are spread evenly over the entire sky. This finding cast doubt on the prevailing wisdom, which held that bursts came from sources within the Milky Way; if they did, the shape of our galaxy, or Earth's off-center position within it, should have caused them to bunch up in certain areas of the sky. The uniform distribution led most astronomers to conclude that the instruments were picking up some kind of event happening through-

A (Very) Warm Afterglow

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

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



out the universe. Unfortunately, gamma rays alone did not provide enough information to settle the question for sure. Researchers would need to detect radiation from the bursts at other wavelengths. Visible light, for example, could reveal the galaxies in which the bursts took place, allowing their distances to be measured. Attempts were made to detect these burst counterparts, but they proved fruitless.

Overview
Gamma-Ray Bursts

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

A Burst of Progress

THE FIELD TOOK a leap forward in 1996 with the advent of the x-ray spacecraft BeppoSAX, built and operated by the Italian Space Agency with the participation of the Netherlands Space Agency. BeppoSAX was the first satellite to localize GRBs precisely and to discover their x-ray "afterglows." The afterglow appears when the gamma-ray signal disappears. It persists for days to months, diminishing with time and degrading from x-rays into less potent radiation, including visible light and radio waves. Although BeppoSAX detected afterglows for only long bursts—eight more years would pass before counterparts of short bursts would be discovered—it made follow-up observations possible at last. Given the positional information from BeppoSAX, optical and radio telescopes were able to identify the galaxies in which the GRBs took place. Nearly all lie billions of light-years away, meaning that the bursts must be enormously powerful. Extreme energies, in turn, call for extreme causes, and researchers began to associate GRBs with the most extreme objects they knew of: black holes.

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

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

The BeppoSAX and follow-up observations have transformed astronomers' view of GRBs. The old concept of a sudden release of energy concentrated in a few brief seconds has been discarded. Indeed, even the term "afterglow" is now recognized as misleading: the energy radiated during both phases is comparable. The spectrum of the after-

MARK A. GARLICK (*preceding pages*); ENRICO COSTA *Institute of Space Astrophysics and Cosmic Physics*, CNR AND THE BEPPOSAX TEAM (*left*); PAUL J. GROOT *University of Amsterdam* (*right*) glow is characteristic of electrons moving in a magnetic field at or very close to the speed of light.

GRB990123—the January 1999 burst—was instrumental in demonstrating the immense power of the bursts. If the burst radiated its energy equally in all directions, it must have had a luminosity of a few times 10^{45} watts, which is 10^{19} times as bright as our sun. In comparison, the photonic emissions of a supernova explosion are spread out over several weeks, with a luminosity that is only a tiny fraction of that of a GRB. Even quasars, which are famously brilliant, give off only about 10^{40} watts.

If the burst beamed its energy in particular directions rather than in all directions, however, the luminosity estimate would be lower. Evidence for beaming comes from the way the afterglow of GRB990123, among others, dimmed over time. Two days into the burst, the rate of dimming increased suddenly, which would happen naturally if the observed radiation came from a narrow jet of material moving at close to the speed of light. Because of a relativistic effect, the observer sees more and more of the jet as it slows down. At some point, there is no more to be seen, and the apparent brightness begins to fall off more rapidly [*see box on next page*].

For GRB990123 and several other bursts, the inferred jet-opening angle is a few degrees. Only if the jet is aimed along our line of sight do we see the burst. This beaming effect reduces the overall energy emitted by the burst approximately in proportion to the square of the jet angle. For example, if the jet subtends 10 degrees, it covers about one 500th of the sky, so the energy requirement goes down by a factor of 500; moreover, for every GRB that is observed, another 499 GRBs go unseen. Even after taking beaming into account, however, the luminosity of GRB990123 was still an impressive 10^{43} watts.

GRB-Supernova Connection

ONE OF THE MOST interesting discoveries has been the connection between GRBs and supernovae. When telescopes went to look at GRB980425, they also found a supernova, called SN1998bw, that had exploded at approximately the same time as the burst. The probability of a chance coincidence was one in 10,000. A firmer case is the association of GRB030329 with SN2003dh. This GRB was localized by NASA's second High Energy Transient Explorer satellite (HETE-2), launched in October 2000. Ground-based observations revealed the broad spectroscopic features of a supernova, basically identical to those of SN1998bw, 10 days after the GRB. The best case by far is GRB060218, which is tied quite nicely to SN2006aj. This GRB was discovered by NASA's Swift satellite, launched in November 2004. Ground-based telescopes were intensely scrutinizing the fading afterglow when the supernova appeared, three days after the GRB.

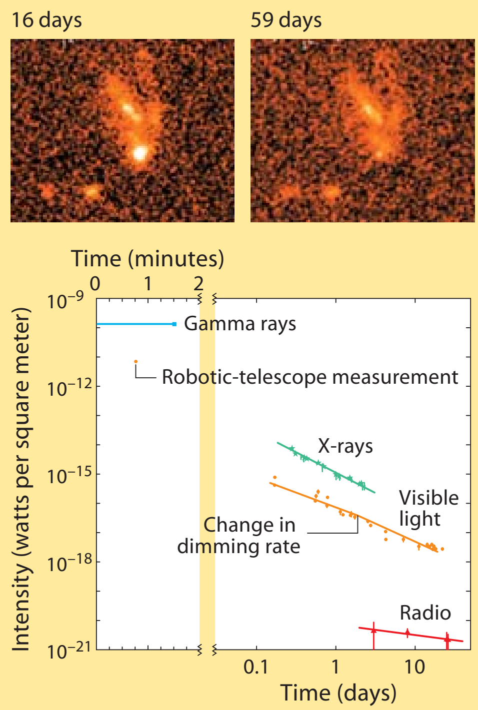
Of the three cases mentioned above, GRB030329 comes closest to being a normal long GRB; GRB980425 and GRB060218 are unusual in that they are underluminous, of long duration and predominantly x-ray events. Also, these two bursts occurred in, by GRB standards, relatively nearby galaxies. The two have long spectral lags, meaning that the high- and low-energy gamma-ray pulses arrive several seconds apart. These bursts may be best described as "x-ray flashes," which will be explained later.

In addition to GRB030329, there is strong evidence that other normal long GRBs are associated with supernovae. GRB970228 was the first BeppoSAX GRB for which an optical afterglow was discovered. At 30 days after the burst, a bump in its optical light curve appeared that looked a lot like a supernova.

A link between GRBs and supernovae has also been suggested by the detection of metals, most notably iron, in the x-ray spectra of several bursts. Iron atoms are known to be synthesized and dumped into interstellar space by supernovae explosions. If these atoms are stripped of their electrons and later hook up with them again, they give off light at distinctive wavelengths, referred to as emission lines. Early detections of such lines by BeppoSAX and the Japanese x-ray satellite ASCA have

Fading Away

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

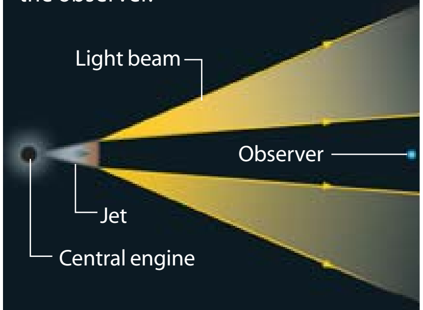


ANDREW FRUCHTER *Space Telescope Science Institute* AND NASA; CORNELIA BLIK (*graph*)

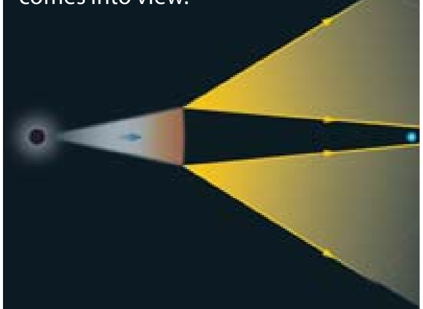
Beam Lines

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

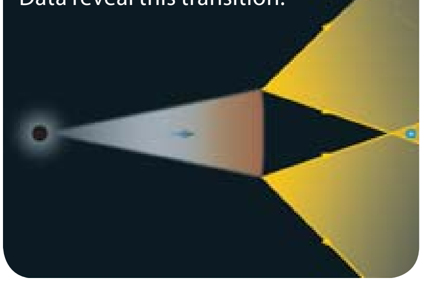
1 Moving at close to the speed of light, the jet emits light in narrow beams. Some beams bypass the observer.



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



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



JUAN VELASCO

been followed up with more solid measurements. Notably, NASA's Chandra X-ray Observatory detected iron lines in GRB991216, which yielded a direct distance measurement of the GRB. The figure agreed with the estimated distance of the burst's host galaxy. And in the shell of gas around GRB011211, the European Space Agency's X-ray Multi-Mirror Satellite found evidence of emission lines from silicon, sulfur, argon and other elements commonly released by supernovae.

A connection between GRBs and supernovae is now generally accepted by astronomers. Because GRBs are much rarer than supernovae—every day a couple of GRBs go off somewhere in the universe, as opposed to hundreds of thousands of supernovae—not every supernova can be associated with a burst. Perhaps jetting inside a supernova is common, and, in some small fraction of cases, relativistic jets escape from the supernova, and, in some small fraction of *those* cases, one of the jets is directed toward us, which allows us to observe a GRB. Also, if the jet is pointed just slightly away from us, then we may observe a lower-energy event, with more x-rays than gamma rays.

Great Balls of Fire

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

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

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

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

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

Many celestial phenomena involve a hole-disk combination. What distinguishes this particular type of system is the sheer mass of the disk (which allows for a gargantuan release of energy) and the lack of a companion star to resupply the disk (which means that the energy release is, more or less, a one-shot event). The black hole and disk have two large reservoirs of energy: the gravitational energy of the disk and the rotational energy of the hole. Exactly how these are converted into a fireball is not fully understood. It is possible that a magnetic field, 10^{15} times more intense than Earth's magnetic field, builds up during the formation of the disk. In so doing, it heats the disk to such high temperatures that it unleashes a fireball of photons, neutrinos and plasma. The fireball is funneled into a pair of narrow jets that flow out along the rotational axis.

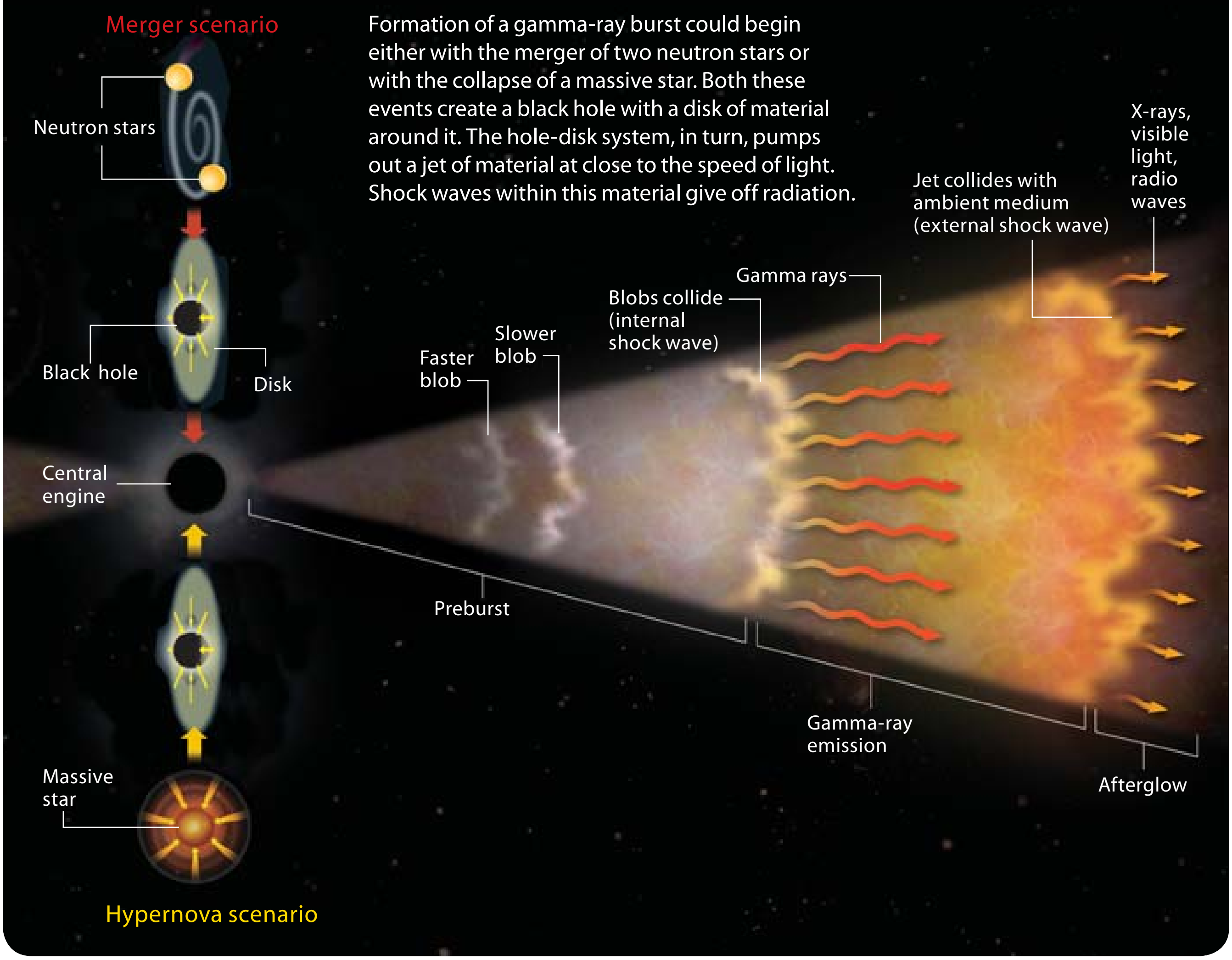
In addition to collapsar and compact-object merger models, it should be noted that there are other models for the central engine of a GRB. One involves the extraction of energy from an electrically charged black hole. In this scenario, both the prompt and afterglow GRB emissions are a result of the fireball sweeping up the external medium.

There is quite a bit of evidence to support the hypothesis that collapsars account for the long GRBs. In particular, the association of long GRBs with supernovae is a point in favor of collapsars, which, after all, are essentially large supernovae. Furthermore, long GRBs are usually found just where collapsars would be expected to occur—namely, in areas of recent star formation within galaxies. A massive star blows up fairly soon (a few million years) after it is born, so its deathbed is close to its birthplace.

The evidence is growing that compact-star coalescence accounts for the short-duration GRBs. This mechanism is not expected to produce a supernova, and, indeed, an association between short GRBs and supernovae has not been found. Also, the decay of the orbit of a pair of compact stars is a process that occurs on a range of timescales, from tens of millions to billions of years. In the former case, the merger will occur close to where the stars in the compact pair were born. In the latter case, the pair will drift around its host galaxy, and so the final coalescence is unlikely to have an association with any star-forming region. Such a mixed association of short GRBs with star

Bursting Out

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



JUAN VELASCO

formation is exactly what was found after Swift and HETE-2 discovered and localized the x-ray afterglows of several short bursts in 2005.

We still do not completely understand the differences between long and short GRBs. For example, the recent burst GRB060614 was a bright, well-observed, nearby event that does not fit cleanly into either category.

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

Blasts from the Past

ONE OUTSTANDING question concerns the dark, or "ghost," GRBs. Of the roughly 200 GRBs that have been localized and studied at wavelengths other than gamma rays, about 90 percent have been seen in x-rays. In contrast, only about 50 percent have been seen in visible light. Why do some bursts fail to shine in visible light?

There are several effects that can make a burst dark. One explanation is that these GRBs lie in regions of star formation, which tend to be filled with dust. Dust blocks visible light but not x-rays. Another intriguing possibility is that some of the ghosts are GRBs that happen to be very far away. The relevant wavelengths of light produced by these bursts would be absorbed by intergalactic gas. To test this hypothesis, measurement of the distance via x-ray or infrared spectra will be crucial. A third possibility is that ghosts are optically faint by nature.

High-sensitivity optical and radio investigations have identified the probable host galaxies of some dark GRBs. Most of them are at moderate distances, favoring—for these events—the dust explanation. But one of them, pinpointed by Swift, is at high redshift in the dark universe region.

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

There are three possible explanations for the x-ray flashes. One is that the fireball is loaded with a relatively large amount of baryonic matter such as protons, making for a "dirty fireball." These particles increase the inertia of the fireball, so that it moves more slowly and is less able to boost photons into the gamma-ray range. Alternatively, the x-ray flashes could be typical GRBs with jets that are pointing just out of our view, so that only the less collimated and less energetic x-rays reach us. A third possibility is that at least one of the x-ray flashes appears to be associated with a less extreme supernova explosion than is usual for normal long GRBs. There is speculation that, in this case, a neutron star, not a black hole, was formed in the supernova.

The next step for GRB astronomy is to accumulate observations of hundreds of bursts of all varieties to flesh out the

THE AUTHORS

NEIL GEHRELS, LUIGI PIRO and *PETER J. T. LEONARD* bring both observation and theory to the study of gamma-ray bursts. Gehrels and Piro are primarily observers—they were the lead scientists, respectively, of the Compton Gamma Ray Observatory and the BeppoSAX satellite. Leonard is a theorist, and like most theorists, he used to think it unlikely that the bursts were bright enough to be seen across the vastness of intergalactic space. "I have to admit that the GRBs really had me fooled," he says. Gehrels is chief of the Astroparticle Physics Laboratory at the NASA Goddard Space Flight Center and lead scientist of the Swift satellite. Piro is director of research at the Institute of Space Astrophysics and Cosmic Physics of the INAF (National Institute of Astrophysics) in Rome. Leonard works for ADNET Systems, in support of missions at Goddard.

The Destinies of Massive Stars

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

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

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

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

Classes of Gamma-Ray Bursts

Burst Class (subclass)	Percentage of All Bursts	Typical Duration of Initial Emission (seconds)	Initial Gamma-Ray Emission	Afterglow X-ray Emission	Afterglow Visible Emission	Hypothetical Central Engine	Explanation for Peculiar Properties
Long (normal)	25	20	Y	Y	Y	Energetic explosion of massive star	Not applicable
Long (ghosts or dark)	30	20	Y	Y	N	Energetic explosion of massive star	Extremely distant, obscured by dust, or intrinsically faint
Long (x-ray-rich or x-ray flashes)	25	30	N	Y	Y	Energetic explosion of massive star	Jet weighed down by extra particles, misaligned jet, or neutron star formed instead of black hole
Short	20	0.3	Y	Y	Y	Merger of pair of compact objects	Does not occur in a star-forming region, so ambient gas is less dense and external shocks are weaker

data on burst, afterglow and host-galaxy characteristics. This effort is being spearheaded by the Swift satellite, thanks to its multiwavelength capabilities and its ability to quickly and autonomously reorient itself to better observe a burst with its high-resolution instruments. Swift's sensitivity to short-duration bursts has been a major factor in understanding this poorly studied class.

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

Other missions, though not designed solely for GRB discovery, will also contribute. The International Gamma-Ray Astrophysics Laboratory, launched on October 17, 2002, is detecting more than 10 GRBs a year. The proposed Energetic X-ray Imaging Survey Telescope will have a sensitive gamma-ray instrument capable of detecting thousands of GRBs. The Explorer of Diffuse Emission and GRB Explosions (EDGE) is proposed to observe GRBs as cosmic beacons—to study the early stages of the universe and its evolution over time.

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

MORE TO EXPLORE

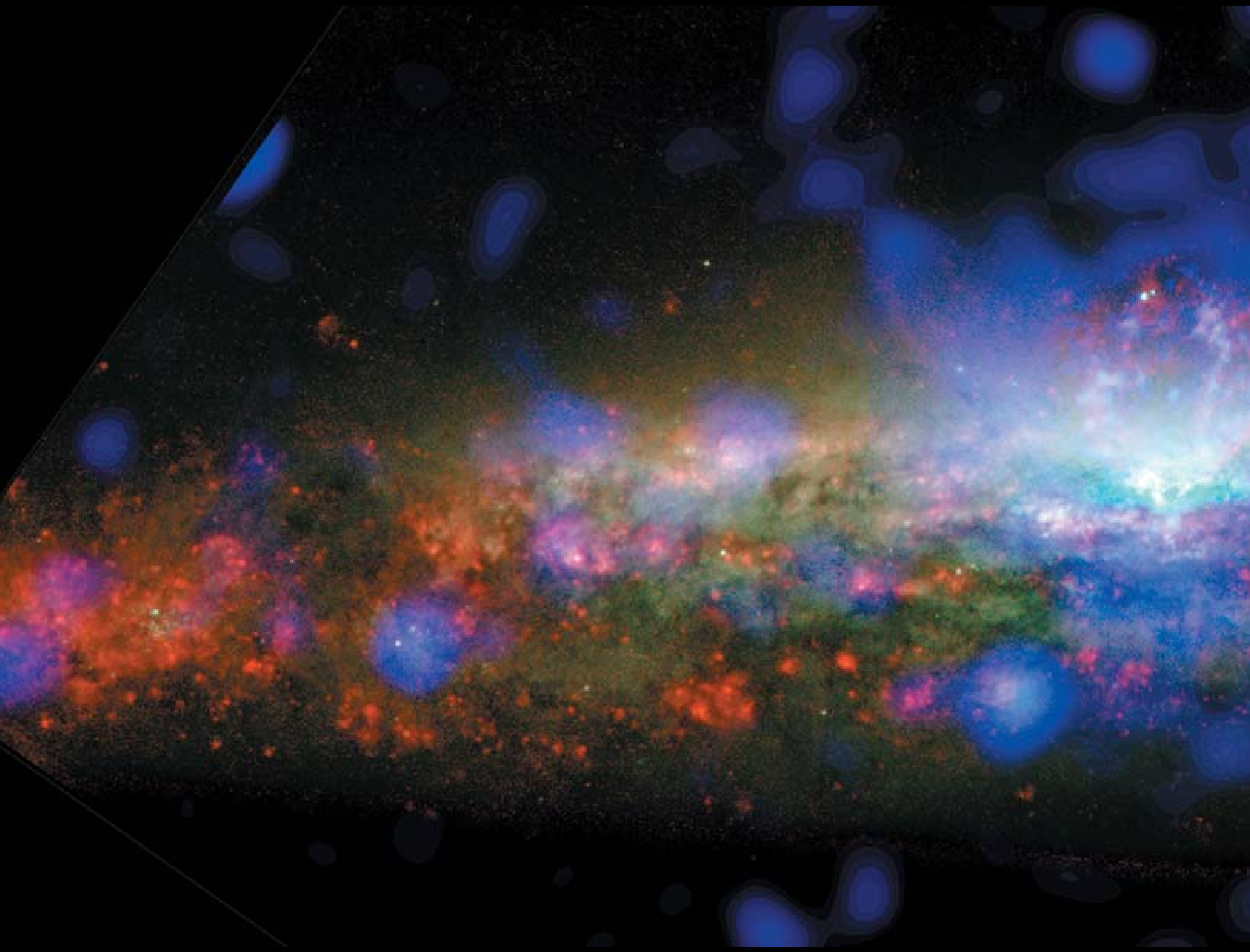
Observation of X-ray Lines from a Gamma-Ray Burst (GRB991216): Evidence of Moving Ejecta from the Progenitor. Luigi Piro et al. in *Science*, Vol. 290, pages 955–958; November 3, 2000. Available online at **arxiv.org/abs/astro-ph/0011337**

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

A Short Gamma-Ray Burst Apparently Associated with an Elliptical Galaxy at Redshift z = 0.225. Neil Gehrels et al. in *Nature*, Vol. 437, pages 851–854; October 6, 2005.

The Association of GRB 060218 with a Supernova and the Evolution of the Shock Wave. Sergio Campana et al. in *Nature*, Vol. 442, pages 1008–1010; August 31, 2006.

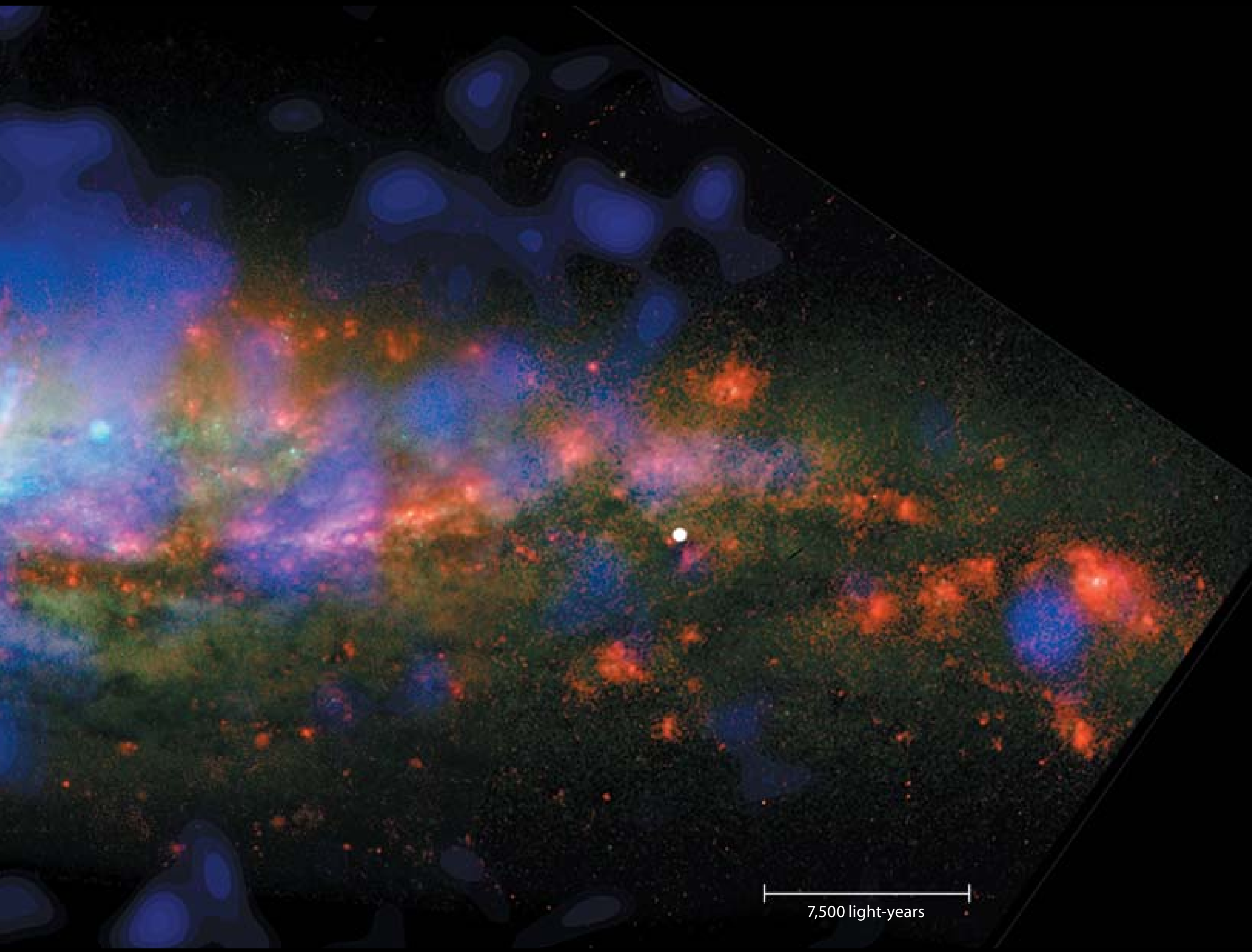
The Supernova–Gamma-Ray-Burst Connection. Stan Woosley and Josh Bloom in *Annual Reviews of Astronomy and Astrophysics*, Vol. 44, pages 507–556; 2006.



the galactic ODD COUPLE

Why do giant black holes and stellar baby booms, two phenomena with little in common, so often go together?

By Kimberly Weaver



Black holes have a bad reputation. In many ways, it is deserved. They are the most efficient engines of destruction known to humanity. Their intense gravity is a one-way ticket to oblivion for anything that strays too close; inside them is undiscovered country from whose bourn no traveler returns. We see them only because the victims do not go quietly to their doom. Material spiraling into a black hole can heat up to millions of degrees and glow brightly. Some of its kinetic energy and momentum may be transferred to a jet of particles flowing outward at close to the speed of light. Black holes of various sizes take the rap for fusillades of radiation and plasma that astronomers observe all over the cosmos.

Yet black holes are not all-powerful. Even those found at the centers of many galaxies, supermassive black holes—whose very name connotes a voracious monster that rules its galactic roost—are minuscule by cosmic standards. They typically count for less than a percent of their galaxy's mass, and

WRETCHED GALAXY NGC 3079 is among those wracked by both of the two most powerful phenomena in the universe: an outburst of star formation and an actively feeding supermassive black hole. As a result, a cone-shaped bubble of hot gas is bursting out of the center of the galaxy at nearly 1,000 kilometers a second. This image combines Hubble Space Telescope visible-light data (*red* and *green*) and Chandra X-ray Observatory data (*blue*).

their gravity is highly concentrated. Accordingly, astronomers long assumed that supermassive holes, let alone their smaller cousins, would have little effect beyond their immediate neighborhoods. Star formation farther out in the galaxy was thought to march to the beat of a different drummer.

So it has come as a surprise over the past decade that black hole activity and star formation are closely intertwined. In many galaxies where black holes devour material greedily—generating a phenomenon that astronomers call an active galactic nucleus (AGN)—stars form at a precipitous rate in episodes known as starbursts. How can these two seemingly disconnected processes be so intimately related?

Today the AGN-starburst connection is a revolutionary area of research. Beautiful Hubble Space Telescope images are allowing astronomers to pick apart the complex events at the hearts of galaxies, the Chandra X-ray Observatory is peering into places hidden to Hubble, and theorists are trying to make sense of it all. This research bears on some of the most basic questions in astronomy: How did the dark early universe come to light up with billions of stars? Did supermassive black holes need a helping hand to grow to be so big? Could they be agents of creation as well as destruction?

Galaxies on Steroids

BOTH ACTIVE GALACTIC NUCLEI and starbursts are among the most spectacular phenomena in the universe. An AGN is a luminous and compact source of light at the center of a galaxy. Quasars are the most extreme example. Pumping out as much power as a billion to a trillion suns, AGNs can outshine the rest of their host galaxies. The supermassive black holes that are thought to power them pack a million to a billion times the sun's mass inside a region smaller than 1,000 times the sun's diameter. Like a falling rock, material spiraling toward the hole picks up speed and releases energy as it collides with other material. In so doing, it gives off radiation at all wavelengths: radio, infrared, optical, ultraviolet, x-ray, gamma-ray.

Starburst galaxies rival the brilliance of AGNs. They are places where gas condenses into stars at a rate equivalent to producing up to 1,000 suns a year—1,000 times faster than stars currently form in our own galaxy. Some starbursts are confined to comparatively small regions, only hundreds of light-years across, located near the center of a galaxy; others occur on much larger scales, sometimes tens of thousands of light-years across. Starbursts often take place in galaxies that are going through, or have recently undergone, a close encounter or merger with a neighboring galaxy. The tidal forces between the two galaxies disrupt gas and cause it to fall inward, greatly accelerating the normal process by which interstellar clouds collapse and form stars. A starburst typically lasts about 10 million years before running out of gas (literally).

Like AGNs, starburst galaxies shine at a wide range of wavelengths. Much of their power output is simply the light of the stars that have been formed. Starbursts tend to be especially bright sources of infrared radiation, which is produced when interstellar dust absorbs and reradiates starlight. Starbursts also produce a lot of x-rays, which pour forth from massive stars, especially as they die. A massive star goes out with a bang: a supernova explosion, which generates x-rays directly, scatters hot x-ray-emitting debris, and leaves behind a neutron star or a smallish black hole, capable of cannibalizing a companion star and spewing x-rays. The surrounding interstellar gas, heated by all the stellar activity, gives off x-rays, too.

The idea that AGNs are somehow linked to starbursts was not sparked by a single earthshaking discovery but has evolved slowly. It goes back to a time when astronomers were still debating what powered AGNs. Although today nearly all attribute AGNs to supermassive black holes, the situation was not so clear as recently as 20 years ago. Researchers, including Roberto Terlevich of the University of Cambridge and Jorge Melnick of the European Southern Observatory, argued that AGNs were a type of starburst. To the telescopes of the day, a tight knot of young stars and supernova debris would look just like a supermassive black hole.

The Case for a Connection

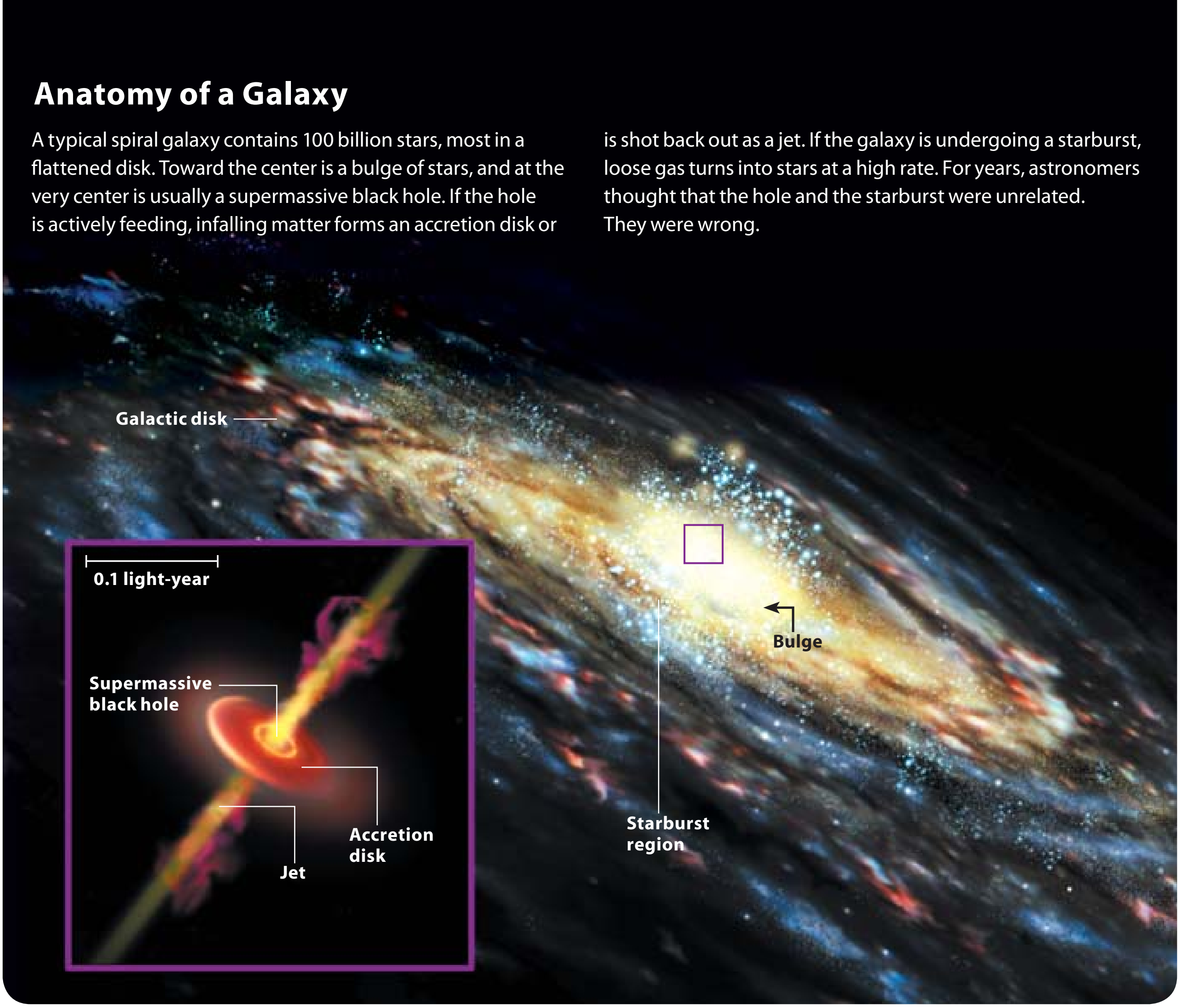
THE NOTION FELL FROM FAVOR only in the late 1980s, as higher-resolution telescopes operating at multiple wavelengths began to reveal just how compact AGNs are: at most a few light-years across and probably a matter of light-minutes across, far too small to encompass a starburst. Even if an entire cluster of stars could fit into such a small space, the stars would rapidly merge together and collapse into a black hole anyway. In addition, AGNs tend to be accompanied by fast-moving jets of material—as a black hole, but not a starburst, would naturally produce.

Although AGNs and starbursts proved to be distinct phe-

Overview
AGNs and Starbursts

- The two most powerful phenomena in galaxies are active galactic nuclei (AGNs) and starbursts. The former are intense, concentrated sources of light—probably matter falling into a supermassive black hole. (Quasars are the best-known example.) Starbursts are galactic fireworks shows during which stars form at a frenetic pace.
- Astronomers used to think that AGNs and starbursts, which are often separated by vast distances, had nothing to do with each other. But they have found that the two phenomena tend to occur hand in hand.
- Does an AGN cause the starburst? Or vice versa? Or are they both caused by some underlying process? The answer will be crucial to understanding the evolution of galaxies.

NASA/CXC/GERALD CECIL *University of North Carolina at Chapel Hill* (*preceding pages*)



nomena, these discussions primed astronomers to accept that they might be related in some way. Several pillars of observational evidence now point to just such a relation. The findings come in a bewildering variety, suggesting that the connection has had a pervasive effect on the universe.

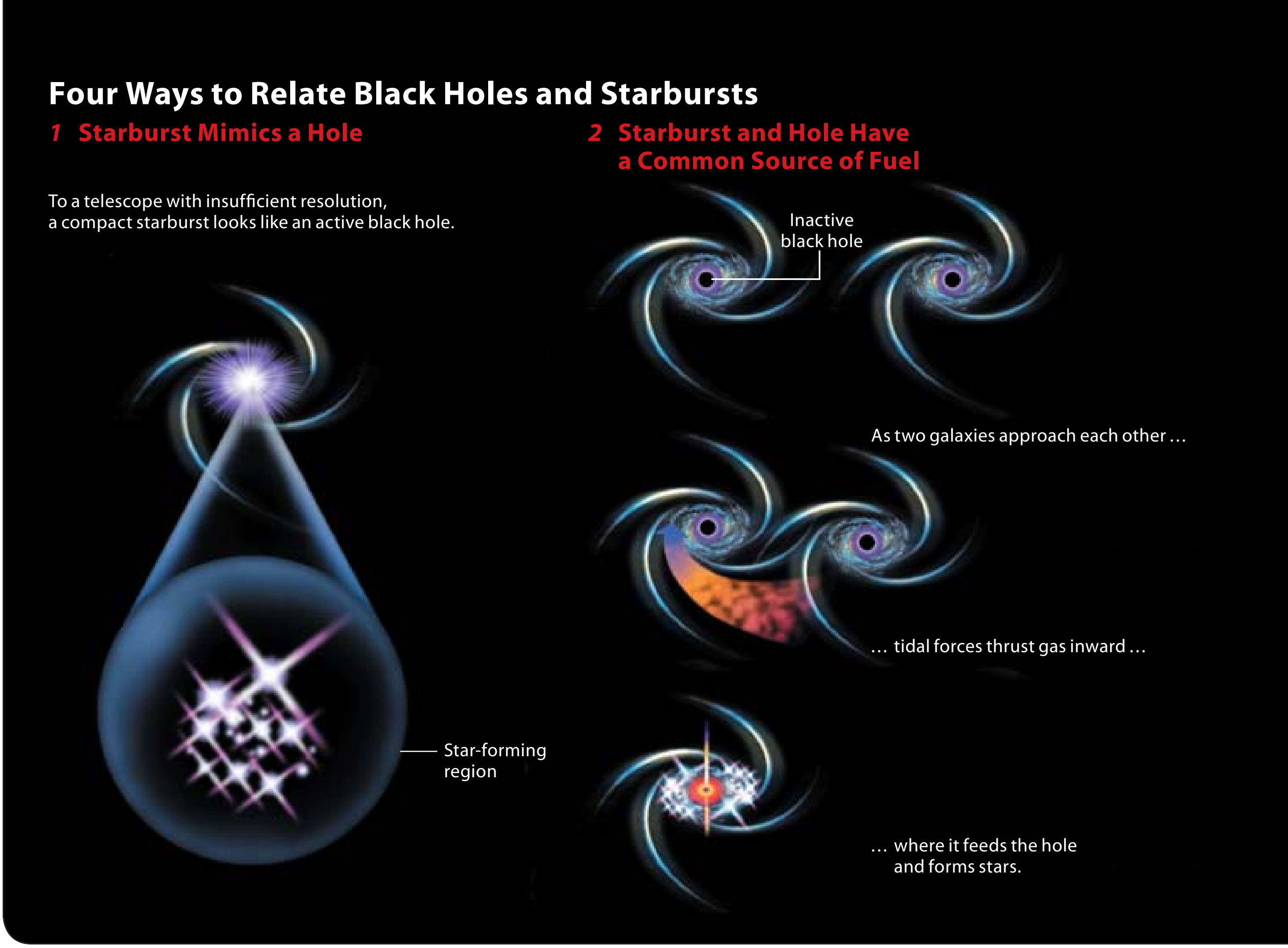
The first piece of evidence is the most direct. Telescopes have seen AGNs alongside starbursts in nearby galaxies. These observations have been tricky to make because galactic cores are filled with gas and dust, obstructing our view. This is where x-ray astronomy comes in. X-rays can penetrate dense gas. Even though current x-ray telescopes lack the resolution of Hubble, they often produce clearer pictures of the dusty centers of galaxies.

A second line of evidence comes from a survey of nearly 23,000 AGNs by Timothy Heckman of Johns Hopkins University and his colleagues. Rather than scrutinizing images of all those galaxies, the researchers inferred the presence of AGNs or starbursts from the strength of particular spectral lines, taking highly ionized oxygen as a sign of an AGN and strong hydrogen absorption as indicative of a starburst. The main conclusion was that galaxies with powerful AGNs had many more young stars than did similar galaxies without AGNs. The more powerful the AGN, the more likely it was that the galaxy had experienced a major starburst not long ago. In short, this study verified that the AGN-starburst connection is not merely anecdotal.

Third, AGN galaxies are not the only ones to be blessed with supermassive black holes. Astronomers have detected them at the centers of inactive galaxies as well. It seems that giant holes are everywhere. Most of the time they lie dormant and invisible; they produce AGNs only when material falls into them at a large and sustained rate. John Kormendy of the University of Texas at Austin, Douglas O. Richstone of the University of Michigan at Ann Arbor and others have demonstrated a correlation between the mass of these holes and the total mass of stars in the galactic centers: the black hole mass is about 0.1 percent of the stellar mass. The same correlation applies to most (though not all) AGN galaxies. Some process, therefore, has linked central black holes to star formation. Lingering discrepancies show that researchers do not fully understand the link.

An AGN-starburst connection might even lurk a mere 24,000 light-years away—at the core of our own galaxy. Rapid motions of stars and gas around the galaxy's center betray the presence of a concentrated mass equal to that of 2.6 million suns. The radio and x-ray emission from this location indicates that the mass is a supermassive black hole—not a truly active hole but one that does feed occasionally. Some

DON DIXON



have hypothesized that it operates like a mini AGN, slurping up surrounding material at one ten-millionth the rate of a true AGN. Although it is not currently accompanied by a starburst, bright clusters of stars do reside nearby. They could be left over from a burst of star formation several million years ago.

Two other forms of evidence come from looking back in time. Observers have noticed that AGNs and star formation were even more closely related when the universe was a tenth of its current age. Back then, two types of galaxies were more common: ultraluminous infrared galaxies (called ULIRGs) and radio galaxies, which appear to be galaxies either in an early stage of formation or in the process of a galaxy merger. Their cores contain huge amounts—billions of solar masses—of cold, dense gas. And they host both AGNs and intense starbursts. The other historical approach concerns distant and luminous AGNs—specifically, quasars. They frequently live in messy galaxies, whose distorted shapes and unusual colors suggest that they are merging and forming stars at a high rate.

A final line of evidence derives from the x-ray background radiation, a lesser-known cousin of the cosmic microwave background radiation. Studies of the background have unveiled a population of AGNs hidden from optical telescopes. This obscuration has a natural explanation: the AGNs were accompanied by starbursts, which choked the galaxies with dust.

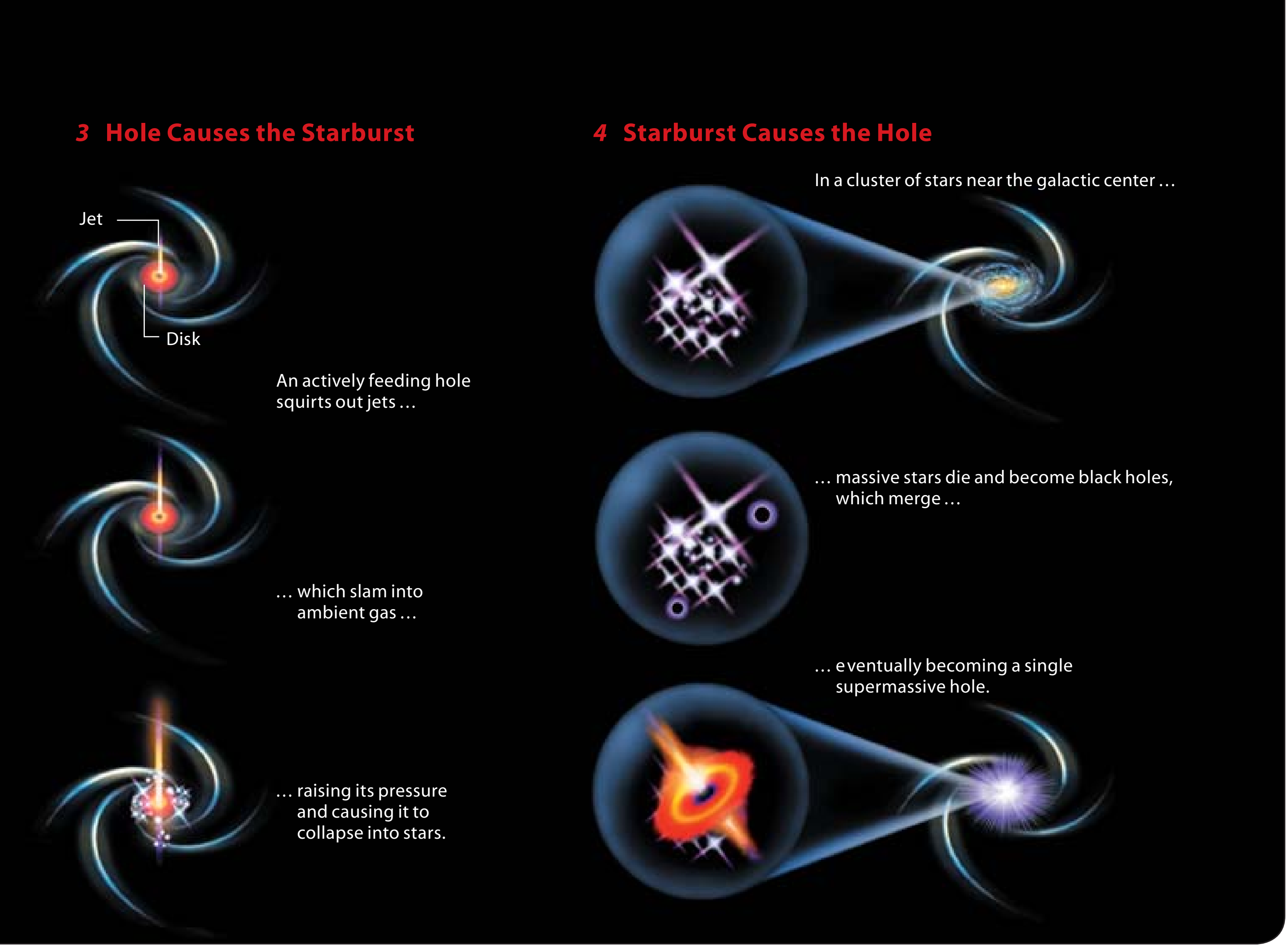
Chicken or Egg?

THE AGN-STARBURST CONNECTION could have come about in four broad ways: the starburst and AGN are one and the same; some third process caused both the AGN and the starburst; the AGN caused the starburst; or the starburst caused the AGN.

The first scenario is a limited version of the older idea that AGNs are simply a type of starburst. Although that idea proved to be wrong for most AGNs, it might work for some of them. Weak AGNs could conceivably be produced by extreme stellar activity rather than a supermassive hole. The activity would occur in such a small region that telescopes might mistake it for a hole. The jury is still out on this possibility.

The second scenario is that the "connection" is merely coincidence. The same processes could set the stage for both starbursts and AGNs. For instance, a galaxy merger could shove gas toward the center of the newly formed entity, inducing a starburst and, by providing fuel for a hole, triggering an AGN. Interestingly, theory predicts that the time it takes for a black hole to grow to supermassive proportions (about 10 million years) is similar to the typical lifetime of a starburst, which is also similar to the time it takes for two galaxies to merge together.

Most researchers, however, have gravitated to the remaining two scenarios, in which AGNs and starbursts are causally related. The third scenario posits that an existing supermas-



sive black hole, contrary to expectation, exerts a strong influence on its host galaxy. Perhaps the hole pulls material toward the galactic center, enabling star formation. Françoise Combes of the Astronomical Observatory of Paris has championed this model. She argues that once a hole is in place, gas naturally flows into the galaxy core, fueling an AGN. As gas collects, it serves as the raw material for a starburst. The theory is quite plausible: many nearby galaxies that host AGNs also contain dusty structures within their cores, which could be material drawn in from outside. On the other hand, not all these structures have the theoretically predicted shape.

Instead of resulting from an inflow of material into the hole, a starburst might be set off by an outflow of energy from the hole. When the supermassive black hole starts to devour material and produce an AGN, shock waves and jets may rip through the galaxy. Gas piles up along shock fronts and condenses into stars. Chandra observations of the Centaurus A galaxy, where the star formation rate is extremely high, suggest that a massive AGN outburst occurred about 10 million years ago. In the outskirts of the galaxy lies a ring of x-ray emission about 25,000 light-years across, which may have resulted from the shock waves of this explosion. The explosion coincided with an episode of star formation, and the x-ray ring overlaps with arcs of young stars.

The black-hole-comes-first scenario has interesting implications. Black holes, rather than stars, may have been the first beacons in the utter blackness of the early universe. Moreover, some astronomers have suggested that the sun was born during a starburst. If this event was triggered by an AGN in the Milky Way, we may owe our existence to a black hole.

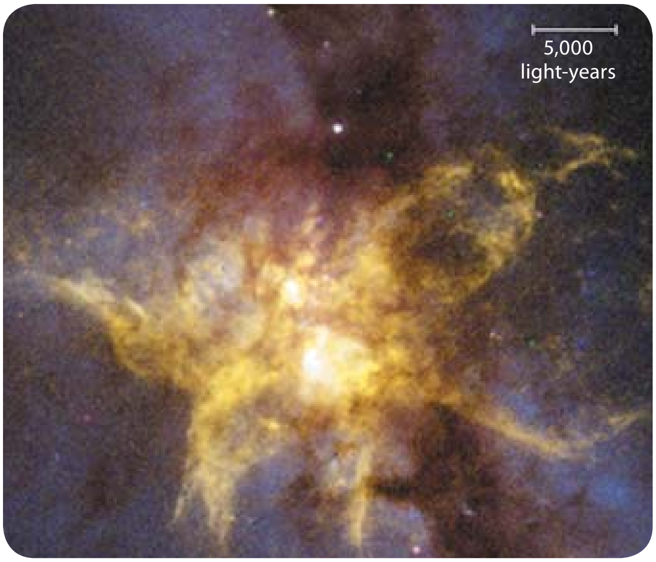
Digging a Hole

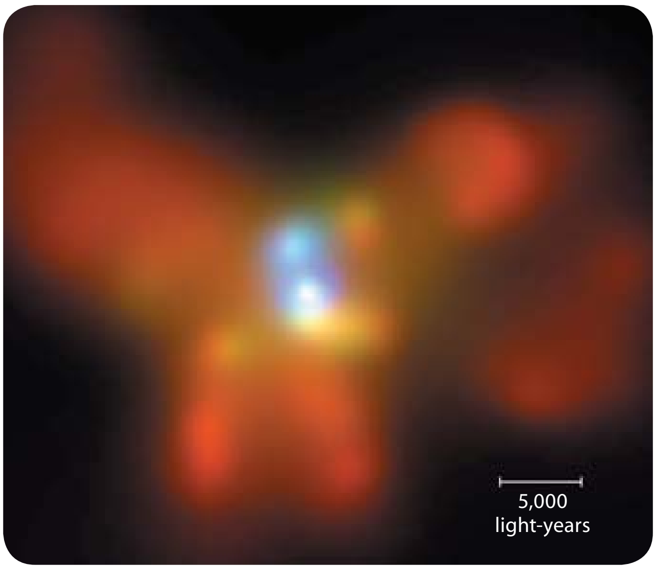
THE STARBURST-COMES-FIRST scenario, though, has the most theoretical and empirical support. The connection can result naturally from normal stellar evolution. A starburst creates dense clusters of stars, within which stellar collisions are common. Massive stars in the cluster quickly die and become neutron stars or stellar-mass black holes, and these bodies agglomerate together. Over tens of millions to hundreds of millions of years, they build up a more massive black hole.

Alternatively, a large black hole could arise from lightweight stars similar to our sun, which do not normally turn into holes. In a dense cluster, these stars could undergo a runaway process of mergers, in which the stars collide and form

THE AUTHOR

KIMBERLY WEAVER is an astrophysicist at the NASA Goddard Space Flight Center and an adjunct professor at Johns Hopkins University. She specializes in x-ray studies of black holes, active galactic nuclei and starburst galaxies. In 1996 Weaver received a Presidential Early Career Award for Scientists and Engineers. She also loves the arts; her hobbies include singing, dancing and acting in community theater.





DOUBLE TROUBLE: As its strange butterfly shape suggests (*top*), NGC 6240 is not one galaxy but a pair of galaxies that recently merged. The system appears to have not one but two supermassive black holes, which show up as distinct sources of x-rays (*blue circles on bottom image*). Diffuse x-ray-emitting gas (*red*) is a sign of rapid star formation. NGC 6240 is a classic example of how holes, starbursts and galaxy mergers occur together.

massive stars, which then join further into megastars a few hundred to a few thousand times as heavy as our sun. Those megastars then collapse to form black holes of similar mass. This process would also take about 100 million years—much less than the lifetime of a galaxy and fast enough to account for the earliest quasars.

No matter how they are created, the black holes would tend to sink into the center of the galaxy. Several could merge to form a supermassive one. This idea has been bolstered by observations of the galaxy NGC 6240, in which a pair of supermassive black holes are circling each other, destined to merge [*see illustration above*]. Supermassive black holes can continue to grow by feasting on surrounding material. Even star clusters that form in distant reaches of a galaxy can contribute mass to the central hole. Those clusters slowly lose kinetic energy and angular momentum because of friction on a galactic scale, caused by dynamical and gravitational interactions with the rest of the galaxy. They spiral inward and eventually get torn apart by tidal forces. Over the course of billions of years, this process could inject into the central black hole a mass equivalent to tens of millions of suns. Disturbances of the galaxy disk, such as an interaction or merger, could likewise pour fuel into the black hole.

Middleweights

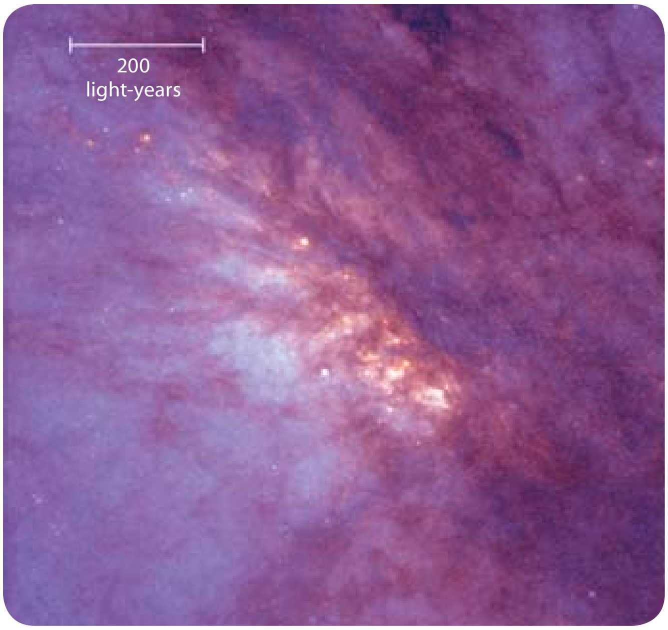
THE STARBURST-COMES-FIRST model predicts an entirely new population of black holes, intermediate between stellar-mass black holes and supermassive ones. Over the past 13 years, circumstantial evidence for these midsize holes has emerged in the form of so-called ultraluminous x-ray sources. Found in several nearby galaxies, these sources emit 10 times to several hundred times as much x-ray power as neutron stars or stellar-mass black holes. They might be neutron stars whose light is beamed in our direction, making them appear abnormally powerful. But evidence is accumulating that they are in fact black holes with a mass of up to several hundred times the mass of the sun.

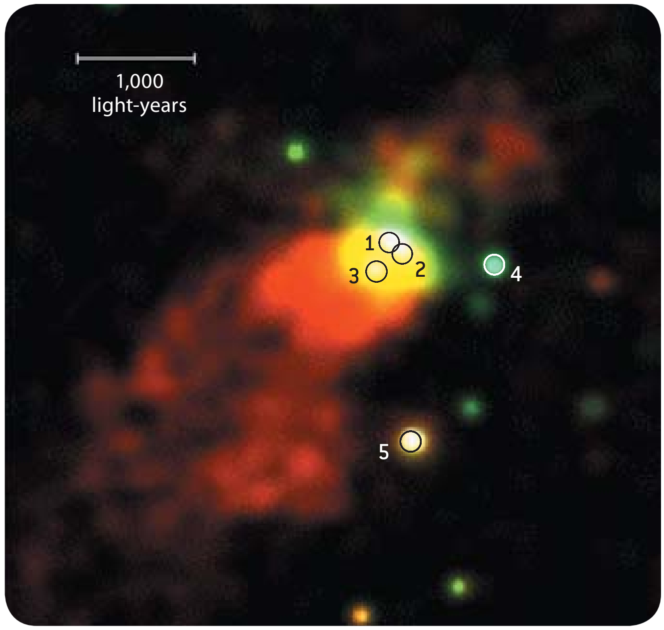
In 2002 two teams of astronomers, one led by Roeland P. van der Marel of the Space Telescope Science Institute in Baltimore and the other by Michael Rich of the University of California, Los Angeles, found hints of intermediate-mass holes at the centers of two dense star clusters, M15 and M31-G1. Stars in these clusters are moving so quickly that it would take bodies of 2,000 and 20,000 solar masses, respectively, to confine them. The "bodies" do not have to be large black holes—they could be a batch of neutron stars or small black holes. But even if that is the case, those objects should eventually merge and produce a large black hole.

Tod Strohmayer and Richard Mushotzky of the NASA Goddard Space Flight Center discovered that one of the ultraluminous sources, located near the center of the starburst galaxy M82, flickers with a period of about 18 seconds. The flickering is too slow and irregular to come from the surface of a neutron star and too intense to come from material in orbit around such a star. If it comes instead from material in orbit around a black hole, the hole could have a mass of several thousand suns. In the spiral galaxy NGC 1313, Jon Miller, then at the Harvard-Smithsonian Center for Astrophysics, and his colleagues found two ultraluminous x-ray sources that are cooler than stellar-mass black holes. Theory predicts that the temperatures near black holes decrease as their mass increases, so the holes in NGC 1313 must be more massive than stellar-mass holes.

These candidate middleweight holes are not located at the centers of their host galaxies, so their relevance to the AGN-starburst connection is not firmly established. But my studies of one nearby starburst galaxy, NGC 253, have provided some direct clues. Before 1995, astronomers believed that the

NASA/ROELAND P. VAN DER MAREL AND JORIS GERSSEN *Space Telescope Science Institute (optical)*; NASA/CXC/STEFANIE KOMASSA *Max Planck Institute for Extraterrestrial Physics (x-ray)*





SMOKING GUN? The central region of galaxy NGC 253 (*left*) suggests that starbursts can build up supermassive black holes. Five x-ray sources (*circles on right image*) are brighter than stellar-mass black holes but dimmer than supermassive ones. They could be medium-size black holes, an intermediate step in the process of creating big holes from mergers of dead stars. Fuzz in the x-ray image is gas associated with star formation.

EUROPEAN SOUTHERN OBSERVATORY (*optical*); NASA/SAO/CXC (*x-ray*)

energetic x-rays from this galaxy originated in the hot gas associated with the starburst. In that year, I found hints of black holes in the x-ray spectrum. It was not until 2001, however, that my colleagues and I obtained an x-ray image of this galaxy with Chandra [*see illustration above*].

We found five ultraluminous x-ray sources within the inner 3,000 light-years of NGC 253. One of them, located exactly at the center of the galaxy, is about 100 times as bright as a neutron star or stellar-mass hole, suggesting that it has a mass equivalent to about 100 suns. It could be a black hole caught in the act of developing into a full-fledged AGN. The sequence of events might go as follows: A starburst takes place near the center of the galaxy. The massive stars thus formed collapse and merge to form lightweight black holes, which then spiral to the galactic center and merge, forming the seed for a supermassive hole. As the starburst winds down, the supermassive hole starts to power an AGN.

Studying how starburst activity affects the fueling and growth of a supermassive hole should offer insight into the birth of the most powerful of all AGNs, quasars. Astronomers have wondered why quasars in the early universe were much more powerful than present-day AGNs. The reason may be simply that the early universe had more frequent episodes of star formation, which triggered more intense AGNs.

To be sure, the situation may be more complicated than a straightforward triggering of one type of activity by the other. Galaxies could cycle between an AGN phase and a starburst phase. When the cycles overlapped, astronomers would see both phenomena together. AGNs and starbursts may even evolve in unison. Current observations are not able to tell whether the AGN comes first, the starburst comes first, or they both occur together. This fascinating question should be answered with the next generation of telescopes.

Observations with the Spitzer Space Telescope, launched by NASA in August 2003, will help illuminate the AGN-starburst connection in some early galaxies. By combining Spitzer's infrared data with visible-light and x-ray data, scientists will be able to determine whether AGNs or starbursts dominate the activity during galaxy formation, which could determine which came first.

The AGN-starburst connection is perhaps the ultimate intergenerational link in the universe. Black holes represent the coalesced embers of bygone stars; starbursts represent the birth of vibrant young stars. It may have taken a partnership of the old and the new to shape galaxies, including ours. SA

MORE TO EXPLORE

Starburst Galaxies: Near and Far. Edited by Linda Tacconi and Dieter Lutz. Springer Verlag, 2001.

X-ray Properties of the Central kpc of AGN and Starbursts: The Latest News from Chandra. Kimberly A. Weaver in *The Central Kiloparsec of Starbursts and AGN: The La Palma Connection.* Edited by J. H. Knapen, J. E. Beckman, I. Shlosman and T. J. Mahoney. ASP Conference Series, Vol. 249; 2001. Available at **arxiv.org/abs/astro-ph/0108481**

Chandra Observations of the Evolving Core of the Starburst Galaxy NGC 253. Kimberly A. Weaver et al. in the *Astrophysical Journal,* Vol. 576, page L19; 2002.

Starburst-AGN Connection from High Redshift to the Present Day. Yoshiaki Taniguchi in *Proceedings of the IAU 8th Asian-Pacific Regional Meeting,* Vol. 1. Edited by Satoru Ikeuchi, John Hearnshaw and Tomoyuki Hanawa. ASP Conference Series, Vol. 289; 2003. **arxiv.org/abs/astro-ph/0211161**

An Introduction to Active Galactic Nuclei. Bradley M. Peterson. Cambridge University Press, 2004.

colossal GALACTIC EXPLOSIONS

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

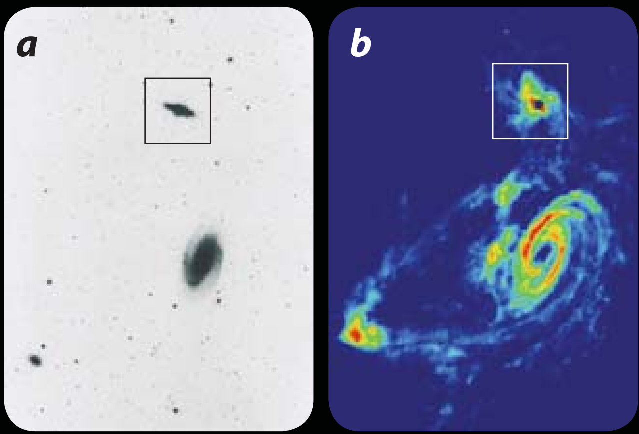
By Sylvain Veilleux, Gerald Cecil and Joss Bland-Hawthorn

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

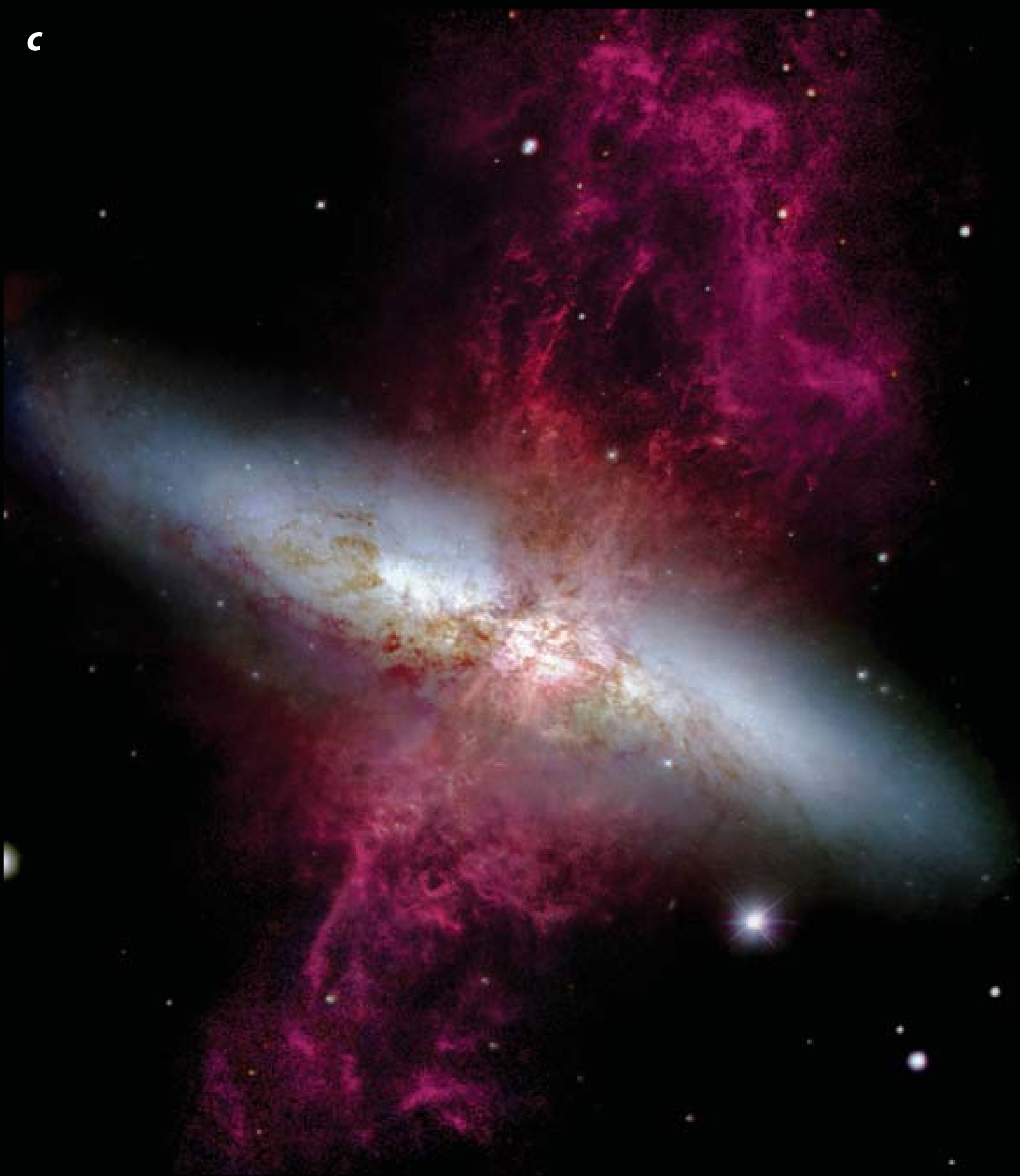
The patterns that this superheated material traces through the interstellar gas and dust surrounding the site of the explosion provide important clues to the nature and history of the powerful forces at work inside the galactic nucleus. Astronomers can now determine what kind of engines drive these dynamos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

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

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



GALAXY M82 (*outlined details in a and b*), about 10 million light-years away from Earth, is distinguished by an outpouring of incandescent gas from the area around its core (*c*). Astronomers have deduced that the upheaval is caused by the rapid formation of stars near the galactic nucleus. The resulting heat and radiation cause dust and gas from the galactic disk to rush into intergalactic space. The galaxy's activity may have been triggered by interaction with its neighbor M81, which cannot be seen in the visible-light image (*a*) but is evident in the radio image of atomic hydrogen gas (*b*).



Superbubble in Space

STARBURST, a sudden pulse of star formation, may be responsible for the activity of NGC 3079 even though the galaxy has a black hole at its center. In these images taken by the Hubble Space Telescope, a close-up view (*inset*) of the area near the nucleus reveals the outlines of an enormous bubble, 3,500 light-years across, that has been blown into the interstellar medium by the heat of the stars forming at the galaxy's center.

times its mass—could power a quasar.

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

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

Some galaxies undergo short episodes of **rapid star formation** in their cores.

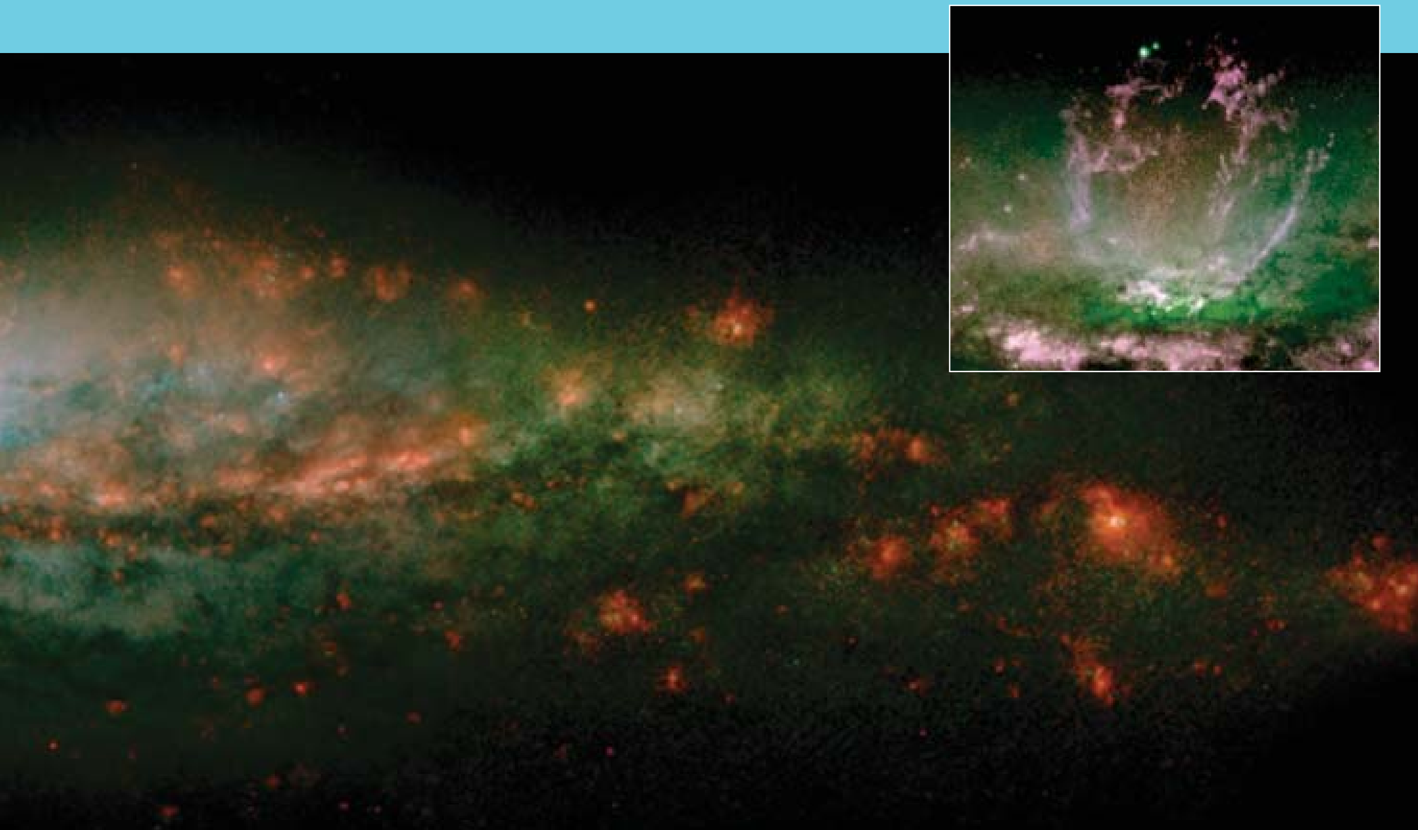
Black holes are not the only engines that drive violent galactic events. Some galaxies apparently undergo short episodes of rapid star formation in their cores: so-called nuclear starbursts. The myriad new stars produce strong stellar winds and, as the stars age, a rash of supernovae. The fast-moving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

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

Identifying the Engine

BOTH THE STARBURST and the black hole explanations appear plausible, but there are important differences between the two that may reveal

COURTESY OF GERALD CECIL



which one is at work in a given galaxy. A black hole can convert as much as 10 percent of the infalling matter to energy. Starbursts, in contrast, rely on nuclear fusion, which can liberate only 0.1 percent of the reacting mass. As a result, they require at least 100 times as much matter, most of which accumulates as unburned fuel. Over the lifetime of a starburst-powered quasar, the total mass accumulated in the nucleus of the galaxy could reach 100 billion times the mass of the sun, equivalent to the mass of all the stars in the Milky Way galaxy.

The more mass near the nucleus, the more rapidly the orbiting stars must move. Ground-based optical observations, which are limited by atmospheric blurring, have not placed tight constraints on the concentration of mass in galactic centers. In 1995, however, radio telescopes discovered an accretion disk with an inner radius of half a light-year spinning rapidly around a mass 20 million times that of the sun at the center of a nearby spiral galaxy called NGC 4258.

Several research groups have mapped the patterns of stellar motions across galactic nuclei using an efficient spectrograph installed on the Hubble telescope by astronauts in 1997. The early discovery that gas in the inner core of galaxy M87 is moving in a manner consistent with a black hole accretion disk demonstrated the promise of such techniques, and subsequent studies have pointed to the presence of black holes at the center of most massive galaxies.

Starbursts and black holes also differ in the spectra of the most energetic photons they produce. Near a black hole, the combination of a strong magnetic field and a dense accretion disk creates a soup of very fast particles that collide with one another and with photons to generate x-rays and gamma rays. A starburst, in contrast, produces most of its high-energy radiation from collisions between supernova ejecta and the surrounding galactic gas and dust.

This impact heats gas to no more than about a billion degrees and so cannot produce any radiation more energetic than x-rays. The large numbers of gamma rays detected from some quasars by the Compton Gamma Ray Observatory imply that black holes are at their centers.

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

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

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

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

For a decade our group used an instrument that combines the advantages of these two methods without the main drawbacks. The Hawaii Imaging Fabry-Perot Interferometer (HIFI) yields detailed spectral information over a large field of view. Named after the turn-of-the-19th-century French inventors Charles Fabry and Alfred Perot, such interferometers have found wide-ranging applications in astronomy. At their heart are two glass plates that are kept perfectly parallel while separated by less than a twentieth of a millimeter. The inner surfaces of the plates are highly reflecting, so light passing through the plates is trapped into repeated reflections. Light of all but a specific wavelength—determined by the precise separation—is attenuated by destructive interference as the light waves bounce back and forth between the plates. By adjusting the separation between the plates, we can produce a series of images that are essentially a grid of spectra obtained by the interferometer at every position over the field of view.

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

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

Sometimes a starburst appears to coexist with a **black hole engine**.

The same regions also emit large amounts of visible and ultraviolet light. We have found, however, that only 10 percent of the light comes from the nucleus. Another 5 percent comes from galaxy-disk gas that has piled up on the expanding edge of the northeast radio lobe. All the rest comes from two fans of high-velocity gas moving outward from the center at high speeds. We have mapped gas moving in excess of 3,000 kilometers a second with the upgraded spectrometer on the Hubble Space Telescope.

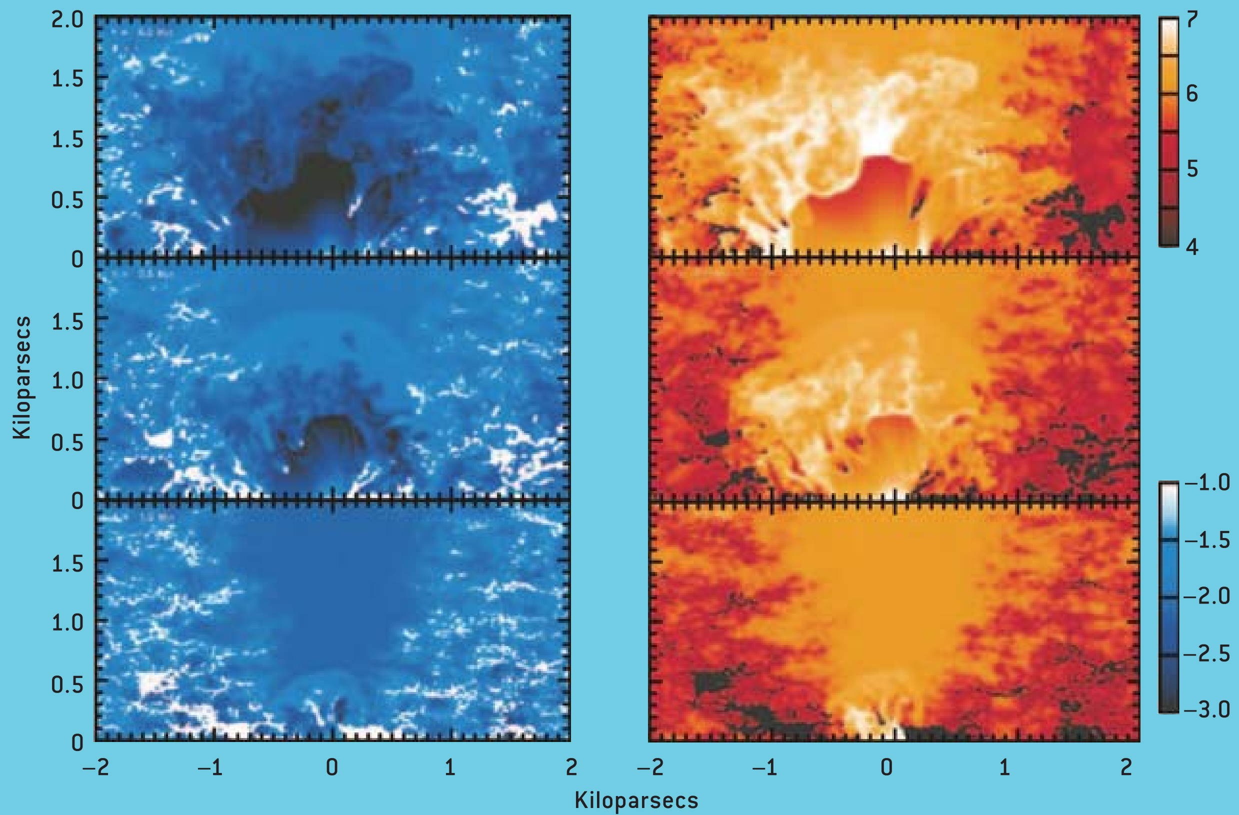
The gas flows outward in two conical regions; it is probably composed of dense filaments of matter that have been swept up by the hot wind from the accretion disk. The axis of the cones of outflowing wind is tilted above the plane of the galaxy but does not point toward the poles.

The effects of the activity within the nucleus reach out several thousand light-years, well beyond the radio lobes. The diffuse interstellar gas exhibits unusually high temperatures, and a large fraction of the atoms have lost one or more electrons and become ionized. At the same time, phenomena in the disk appear to influence the nucleus. Infra-

THE AUTHORS

SYLVAIN VEILLEUX, GERALD CECIL and *JOSS BLAND-HAWTHORN* met while working at observatories in Hawaii and were drawn to collaborate by a shared interest in highly active galaxies. Veilleux, now professor of astronomy at the University of Maryland, received his Ph.D. from the University of California, Santa Cruz. Cecil, professor of astronomy and physics at the University of North Carolina at Chapel Hill, received his doctorate from the University of Hawaii. Aware of the rapidly approaching and premature transition of humanity's energy use away from petroleum, he has in recent years spent more time on energy analysis than astrophysics. Bland-Hawthorn received his Ph.D. in astronomy and astrophysics from the University of Sussex and the Royal Greenwich Observatory in England. He is now head of instrument science at the Anglo-Australian Observatory in Sydney.

Galactic Wind



OUTPOURING OF GAS rapidly becomes turbulent in this computer simulation of an active starburst-driven galaxy. Temperature maps (*right*) show how the hot gas emanating from the nucleus displaces the cooler galactic gas around it. The resulting shock fronts appear as sharp boundaries in maps of gas density (*left*). Time progresses from bottom (one million years old) to top (four million years old). Distances are shown in kiloparsecs, each equivalent to about 3,000 light-years.

COURTESY OF JACKIE COOPER, GEOFFREY BICKNELL AND RALPH S. SUTHERLAND *Australian National University*

red images reveal an elongated bar of stars that extends more than 3,000 light-years from the nucleus. The HIFI velocity measurements suggest that the bar distorts the circular orbits of gas in the disk, funneling material toward the center of the galaxy. This inflow of material may fuel the black hole.

Nearby Active Galaxies

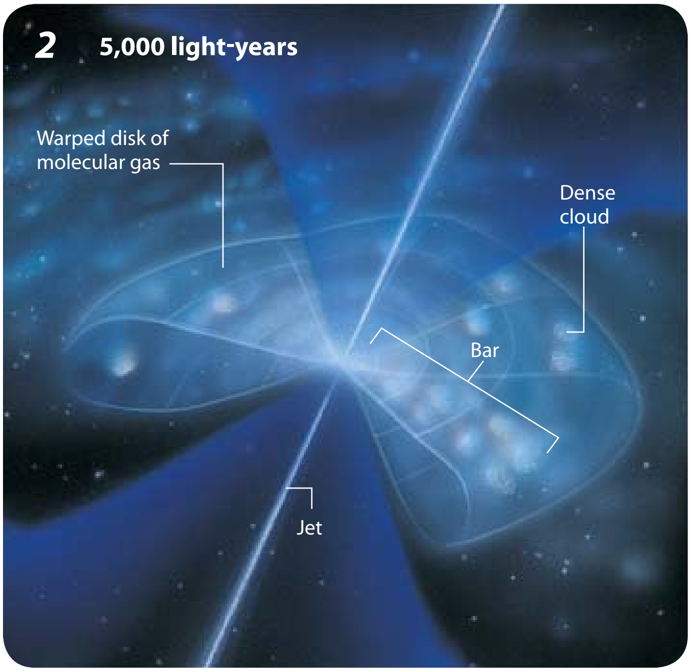
ANOTHER tremendous explosion is occurring in the core of one of our nearest neighbor galaxies, M82, just a few million light-years away. In contrast to NGC 1068, this cataclysm appears to be an archetypal starburst-driven event. Images exposed through a filter that passes the red light of forming hydrogen atoms reveal a web of filaments spraying outward along the galactic poles. Our spectral grids of emission from filaments perpendicular to the galactic disk reveal two main masses of gas, one receding and the other approaching. The difference in velocity between the two increases as the gas moves outward from the core, reaching about 350 kilometers a second at a distance of 3,000 light-years. At a distance of 4,500 light-years from the core, the velocity separation diminishes.

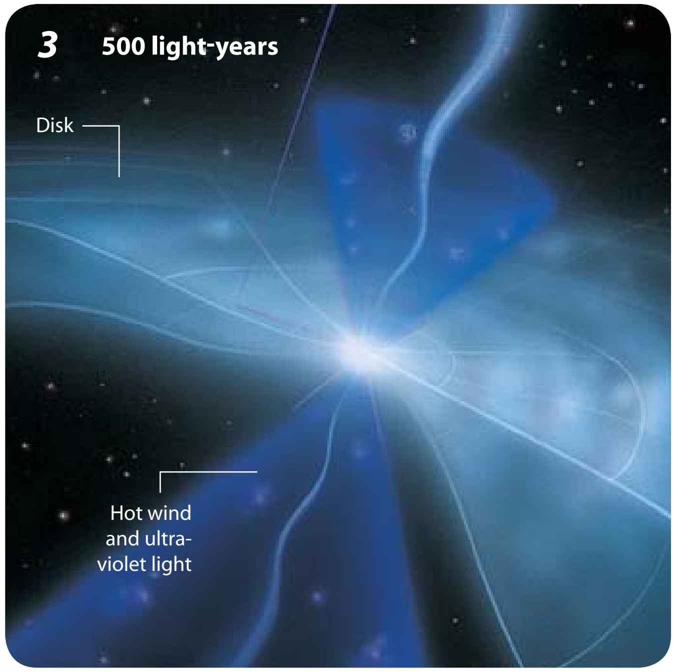
The core of M82 is undergoing an intense burst of star formation, possibly triggered by a recent orbital encounter with its neighbors M81 and NGC 3077. Its infrared luminosity is 30 billion times the total luminosity of the sun, and radio astronomers have identified the remnants of large numbers of supernovae. The filamentary web visible from Earth results from two elongated bubbles oriented roughly perpendicular to the disk of M82 and straddling the nucleus. X-ray observatories in space have detected the hot wind that inflates these bubbles; their foamy appearance probably arises from instabilities in the hot gas as it cools.

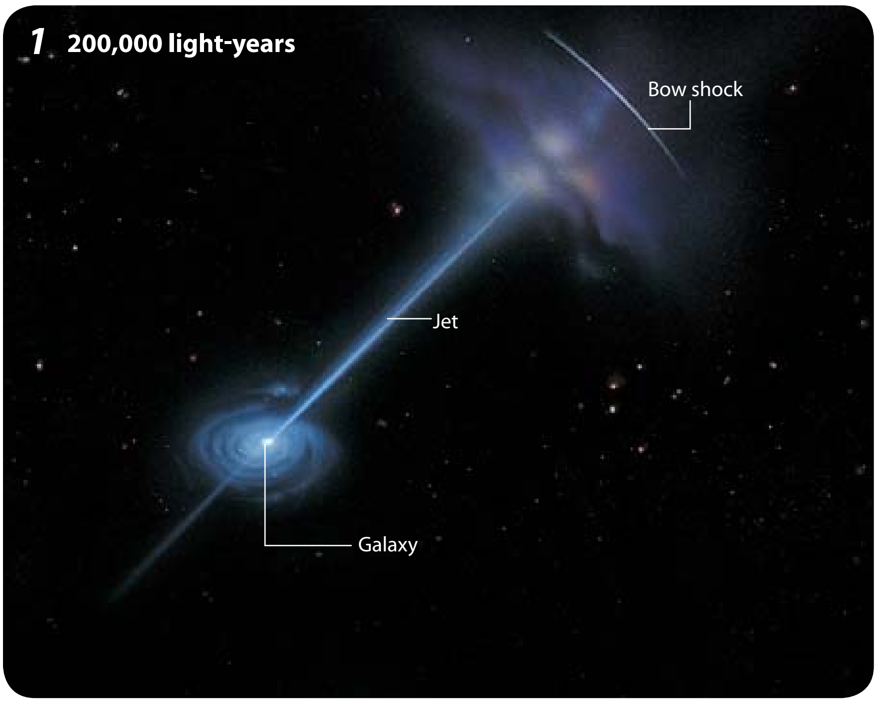
Ambiguous Activity

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

Such ambiguity plagues interpretations of the behavior of the nearby galaxy NGC 3079. This spiral galaxy appears almost edge-on from Earth—an excellent vantage point from which to study the gas expelled from the nucleus. Like galaxy M82, NGC 3079 is anomalously bright in the infrared, and it also







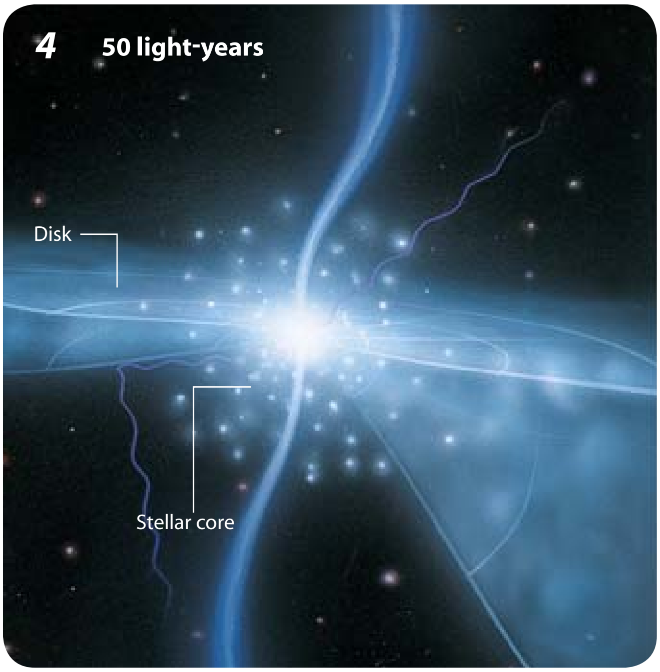
COLOSSAL FORCES at work in the center of an active galaxy can make themselves felt half a million light-years or more away as jets of gas moving at relativistic speeds plow into the intergalactic medium and create enormous shock waves (*1*). Closer to the center of the galaxy (*2, 3*), a dense equatorial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

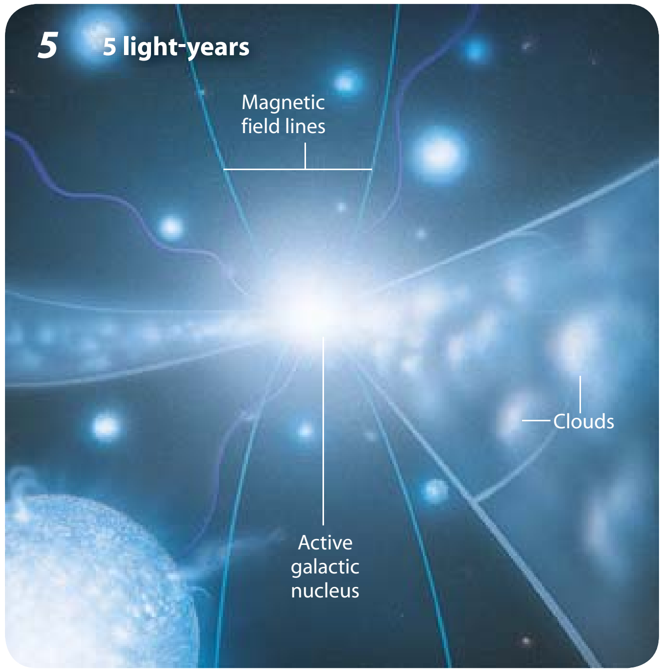
contains a massive disk of molecular gas spanning 8,000 light-years around its core. At the same time, the core is unusually bright at radio wavelengths, and the linear shape of radio-emitting regions near the core suggests a collimated jet outflow. On a larger scale, the radio-emission pattern is complex and extends more than 6,500 light-years from either side of the galactic disk.

Images made in red hydrogen light show a nearly circular ring 3,600 light-years across just east of the nucleus; velocity measurements from the HIFI confirm that the ring marks the edge of a bubble as seen from the side. The bubble resembles an egg with its pointed extremity balanced on the nucleus and its long axis aligned with the galactic pole. There is another bubble on the west side of the nucleus, but most of it is hidden behind the dusty galaxy disk.

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

ALFRED T. KAMAJIAN





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

How Active Galaxies Form

ALTHOUGH ASTRONOMERS now understand the basic principles of operation of the engines that drive active galaxies, many details remain unclear. There is a vigorous debate about the nature of the processes that ignite a starburst or form a central black hole. What is the conveyor belt that transports fuel down to the pointlike nucleus? Most likely, gravitational interactions with gas-rich galaxies redistribute gas in the host galaxy, perhaps by forming a stellar bar such as the one in NGC 1068. Computer simulations appear to indicate that the bar, once formed, may be quite stable. (Indeed, the bar must be stable, because NGC 1068 currently has no close companion.)

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

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

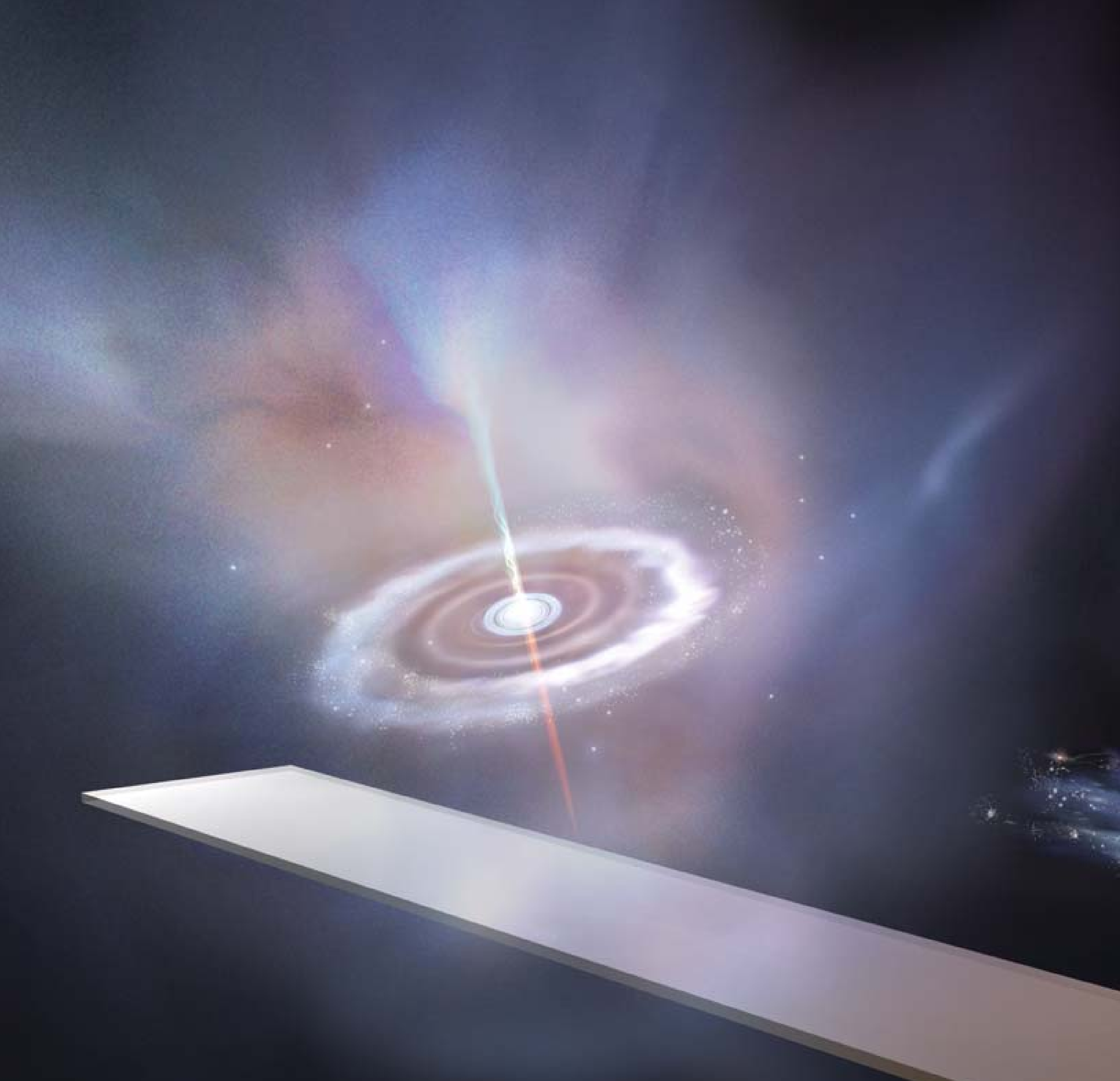
The full impact of this "cosmic bubble bath" over the history of the universe is difficult to assess accurately because we currently know very little of the state of more distant galaxies. Images of distant galaxies taken by the Hubble are clarifying some of these questions. Indeed, as the light that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

MORE TO EXPLORE

Active Galactic Nuclei: From the Central Black Hole to the Galactic Environment. Julian H. Krolik. Princeton University Press, 1998.

An Introduction to Active Galactic Nuclei. Bradley M. Peterson. Cambridge University Press, 2004.

Galactic Winds. Sylvain Veilleux, Gerald Cecil and Joss Bland-Hawthorn in *Annual Review of Astronomy and Astrophysics,* Vol. 43, pages 769–826; 2005.

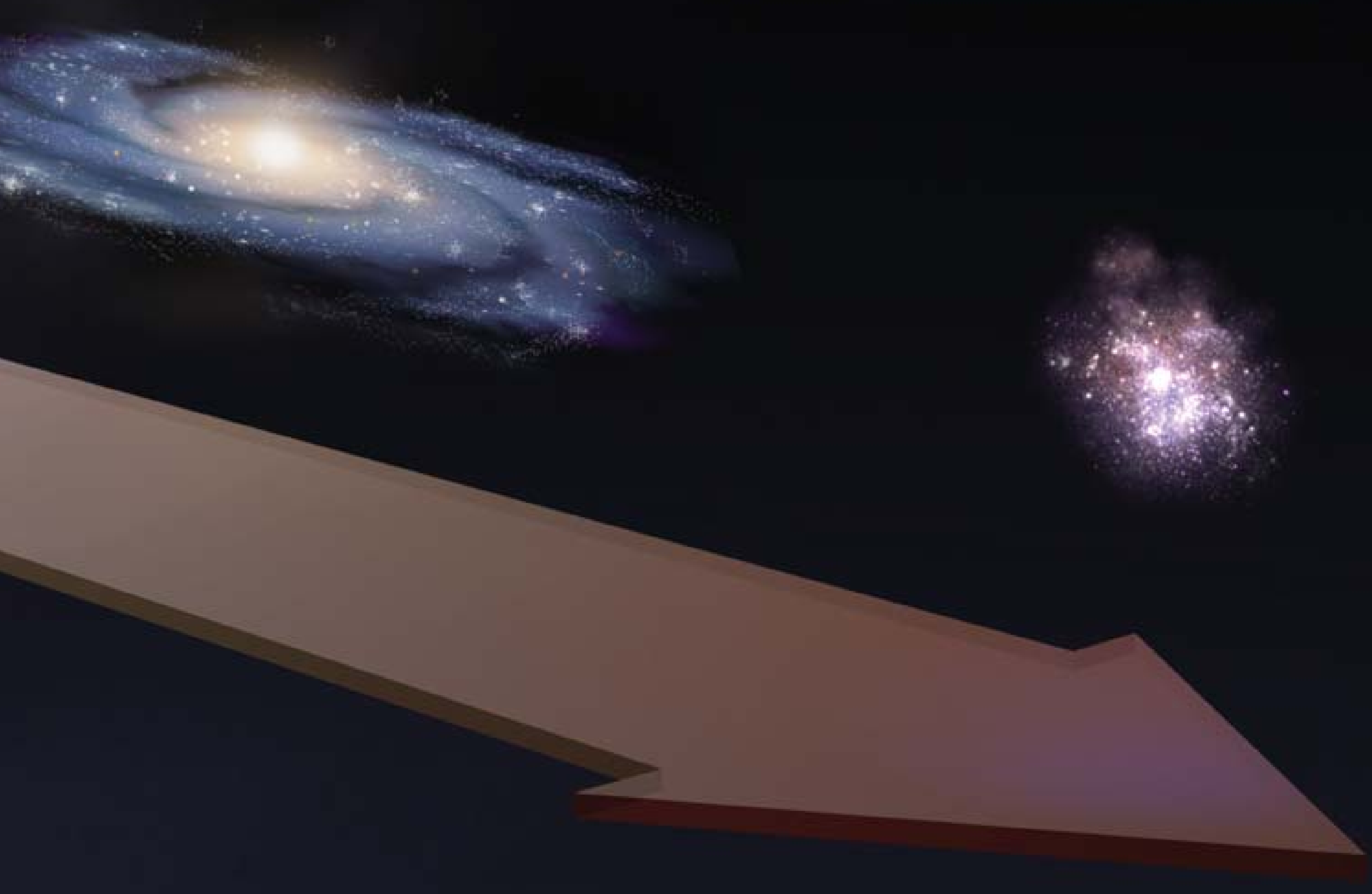


COSMIC DOWNSIZING has occurred over the past 14 billion years as activity has shifted to smaller galaxies. In the first half of the universe's lifetime, giant galaxies gave birth to prodigious numbers of stars and supermassive black holes that powered brilliant quasars (*left*). In the second half, activity in the giant galaxies slowed, but star formation and black hole building continued in medium-size galaxies (*center*). In the future, the main sites of cosmic activity will be dwarf galaxies holding only a few million stars each (*right*).

THE MIDLIFE CRISIS of the cosmos

Although it is not as active as it used to be, the universe is still forming stars and building black holes at an impressive pace

By Amy J. Barger



Until recently, most astronomers believed that the universe had entered a very boring middle age. According to this paradigm, the early history of the universe—that is, until about six billion years after the big bang—was an era of cosmic fireworks: galaxies collided and merged, powerful black holes sucked in huge whirlpools of gas, and stars were born in unrivaled profusion. In the following eight billion years, in contrast, galactic mergers became much less common, the gargantuan black holes went dormant, and star formation slowed to a flicker. Many astronomers were convinced that they were witnessing the end of cosmic history and that the future held nothing but the relentless expansion of a becalmed and senescent universe.

In the past few years, however, new observations have made it clear that the reports of the universe's demise have been greatly exaggerated. With the advent of new space observatories and new instruments on ground-based telescopes, astronomers have detected violent activity occurring in nearby galaxies during the recent past. (The light from more distant galaxies takes longer to reach us, so we observe these structures in an earlier stage of development.) By examining the x-rays emitted by the cores of these relatively close galaxies, researchers have discovered many tremendously massive black holes still devouring the surrounding gas and dust. Furthermore, a more thorough study of the light emitted by galaxies of different ages has shown that the star formation rate has not declined as steeply as once believed.

The emerging consensus is that the early universe was dominated by a small number of giant galaxies containing colossal black holes and prodigious bursts of star formation, whereas the present universe has a more dispersed nature—the creation of stars and the accretion of material into black holes are now occurring in a large number of medium-size and small galaxies. Essentially, we are in the midst of a vast downsizing that is redistributing cosmic activity.

Overview
Middle-Aged Cosmos

- The early history of the universe was a turbulent era marked by galactic collisions, huge bursts of star formation and the creation of extremely massive black holes. The falloff in cosmic activity since then has led many astronomers to believe that the glory days of the universe are long gone.
- In recent years, though, researchers have found powerful black holes still actively consuming gas in many nearby galaxies. New observations also suggest that star formation has not dropped as steeply as once believed.
- The results point to a cosmic downsizing: whereas the early universe was dominated by a relatively small number of giant galaxies, activity in the current universe is dispersed among a large number of smaller galaxies.

Deep-Field Images

TO PIECE TOGETHER the history of the cosmos, astronomers must first make sense of the astounding multitude of objects they observe. Our most sensitive optical views of the universe come from the Hubble Space Telescope. In the Hubble Deep Field studies—10-day exposures of two tiny regions of the sky observed through four different wavelength filters—researchers have found thousands of distant galaxies, with the oldest dating back to about one billion years after the big bang. A more recent study, called the Hubble Ultra Deep Field, has revealed even older galaxies.

Obtaining these deep-field images is only the beginning, however. Astronomers are seeking to understand how the oldest and most distant objects evolved into present-day galaxies. It is somewhat like learning how a human baby grows to be an adult. Connecting the present with the past has become one of the dominant themes of modern astronomy.

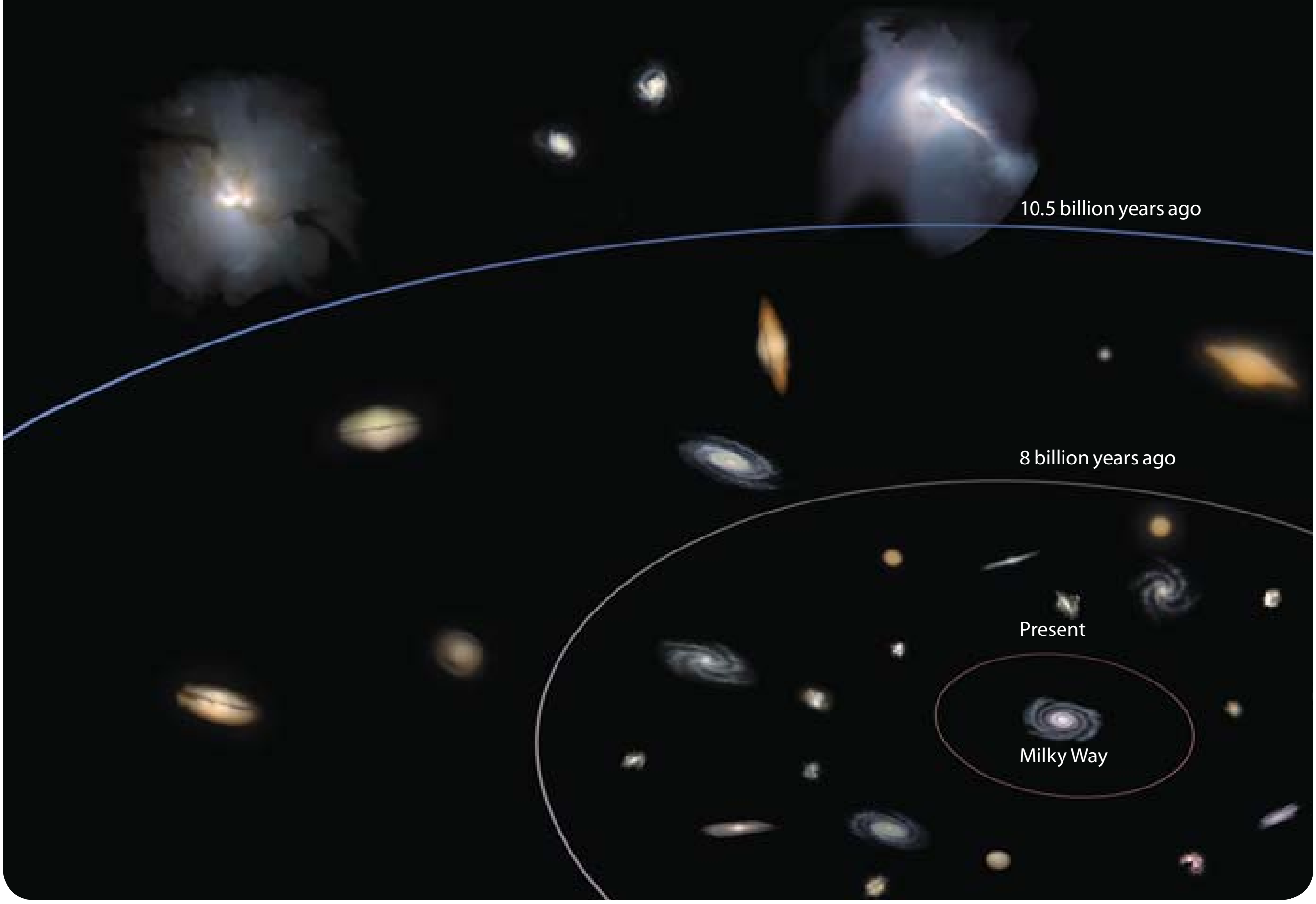
A major step in this direction is to determine the cosmic stratigraphy—which objects are in front and which are more distant—among the thousands of galaxies in a typical deep-field image. The standard way to perform this task is to obtain a spectrum of each galaxy in the image and measure its redshift. Because of the universe's expansion, the light from distant sources has been stretched, shifting its wavelength toward the red end of the spectrum. The more the light is shifted to the red, the farther away the source is and thus the older it is. For example, a redshift of one means that the wavelength has been stretched by 100 percent—that is, to twice its original size. Light from an object with this redshift was emitted about six billion years after the big bang, which is less than half the current age of the universe. In fact, astronomers usually talk in terms of redshift rather than years, because redshift is what we measure directly.

Obtaining redshifts is a practically foolproof technique for reconstructing cosmic history, but in the deepest of the deep-field images it is almost impossible to measure redshifts for all the galaxies. One reason is the sheer number of galaxies in the image, but a more fundamental problem is the intrinsic faintness of some of the galaxies. The light from these dim objects arrives at a trickle of only one photon per minute in each square centimeter. And when

ALFRED T. KAMAJIAN (*preceding pages*)

Evolution of the Universe

As astronomers peer into the depths of space, they also look back in time, because the light from distant objects takes longer to reach us. More than 10.5 billion years ago, tremendous galaxies collided and merged, triggering bursts of star formation and the accretion of gas into supermassive black holes. Between eight billion and 10.5 billion years ago, stars continued to form at a high rate, and black holes continued to grow inside the galactic cores. In more recent times, star formation and black hole activity began to die down in the bigger galaxies; in the present-day universe, most of the star formation takes place in smaller spiral and irregular galaxies.



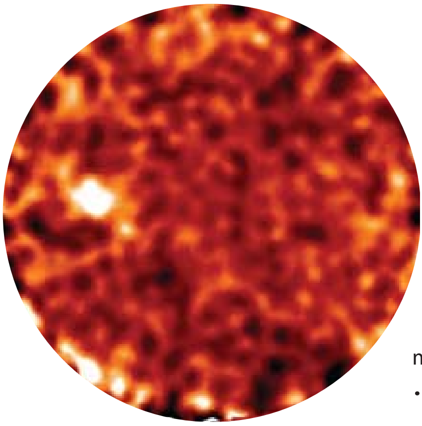
observers take a spectrum of the galaxy, the diffraction grating of the spectrograph disperses the light over a large area on the detector, rendering the signal even fainter at each wavelength.

In the late 1980s a team led by Lennox L. Cowie of the University of Hawaii Institute for Astronomy and Simon J. Lilly, now at the Swiss Federal Institute of Technology in Zurich, developed a novel approach to avoid the need for laborious redshift observations. The researchers observed regions of the sky with filters that selected narrow wavebands in the ultraviolet, green and red parts of the spectrum and then measured how bright the galaxies were in each of the wavebands [*see box on page 63*]. A nearby star-forming galaxy is equally bright in all three wavebands. The intrinsic light from a star-forming galaxy has a sharp cutoff just beyond the ultraviolet waveband, at a wavelength of about 912 angstroms. (The cutoff appears because the neutral hydrogen gas in and around the galaxy absorbs radiation with shorter wavelengths.) Because the light from distant galaxies is shifted to the red, the cutoff moves to longer wavelengths; if the redshift is great enough, the galaxy's light will not appear in the ultraviolet waveband, and if the redshift is greater still, the galaxy will not be visible in the green waveband either.

Thus, Cowie and Lilly could separate star-forming galaxies into broad redshift intervals that roughly indicated their ages. In 1996 Charles C. Steidel of the California Institute of Technology and his collaborators used this technique to isolate hundreds of ancient star-forming galaxies with redshifts of about three, dating from about two billion years after the big bang. The researchers confirmed many of the estimated redshifts by obtaining very deep spectra of the galaxies with the powerful 10-meter Keck telescope on Mauna Kea in Hawaii.

Once the redshifts of the galaxies have been measured, we can begin to reconstruct the history of star formation. We know from observations of nearby galaxies that a small number of high-mass stars and a larger number of low-mass stars usually form at the same time. For every 20 sunlike stars that are born, only one 10-solar-mass star (that is, a star with a mass 10 times as great

ALFRED T. KAMAJIAN



EXTREMELY LUMINOUS GALAXIES in the early universe have been discovered using the Submillimeter Common-User Bolometer Array (SCUBA) on the James Clerk Maxwell Telescope on Mauna Kea in Hawaii. The bright blob on the left is believed to be an ancient dust-enshrouded galaxy that spawned stars at a phenomenal rate, forming the equivalent of more than 1,000 suns every year.

A. J. BARGER, L. L. COWIE, D. B. SANDERS, E. FULTON ET AL. *James Clerk Maxwell Telescope*

as the sun's) is created. High-mass stars emit ultraviolet and blue light, whereas low-mass stars emit yellow and red light. If the redshift of a distant galaxy is known, astronomers can determine the galaxy's intrinsic spectrum (also called the rest-frame spectrum). Then, by measuring the total amount of rest-frame ultraviolet light, researchers can estimate the number of high-mass stars in the galaxy.

Because high-mass stars live for only a few tens of millions of years—a short time by galactic standards—their number closely tracks variations in the galaxy's overall star formation rate. As the pace of star creation slows, the number of high-mass stars declines soon afterward because they die so quickly after they are born. In our own Milky Way, which is quite typical of nearby, massive spiral galaxies, the number of observed high-mass stars indicates that stars are forming at a rate of a few solar masses a year. In high-redshift galaxies, however, the rate of star formation is 10 times as great.

When Cowie and Lilly calculated the star formation rates in all the galaxies they observed, they came to the remarkable conclusion that the universe underwent a veritable baby boom at a redshift of about one. In 1996 Piero Madau, now at the University of California at Santa Cruz, put the technique to work on the Hubble Deep Field North data, which were ideal for this approach because of the very precise intensity measurements in four wavebands. Madau combined his results with those from existing lower-redshift optical observations to refine the estimates of the star formation history of the universe. He inferred that the rate of star formation must have peaked when the universe was about four billion to six billion years old. This result led many astronomers to conclude that the universe's best days were far behind it.

An Absorbing Tale

ALTHOUGH MADAU'S ANALYSIS of star formation history was an important milestone, it was only a small part of the story. Galaxy surveys using optical telescopes cannot detect every source in the early universe. The more distant a galaxy is, the more it suffers from cosmological redshifting, and at high enough redshifts, the galaxy's rest-frame ultraviolet and optical emissions will be stretched into the infrared part of the spectrum. Furthermore, stars tend to reside in very dusty environments because of the detritus from supernova explosions and other processes. The starlight heats up the dust grains, which then reradiate this energy at far-infrared wavelengths. For very distant sources, the light that is absorbed by dust and reradiated into the far-infrared is shifted by the expansion of the universe to submillimeter wavelengths. Therefore, a bright source of submillimeter light is often a sign of intense star formation.

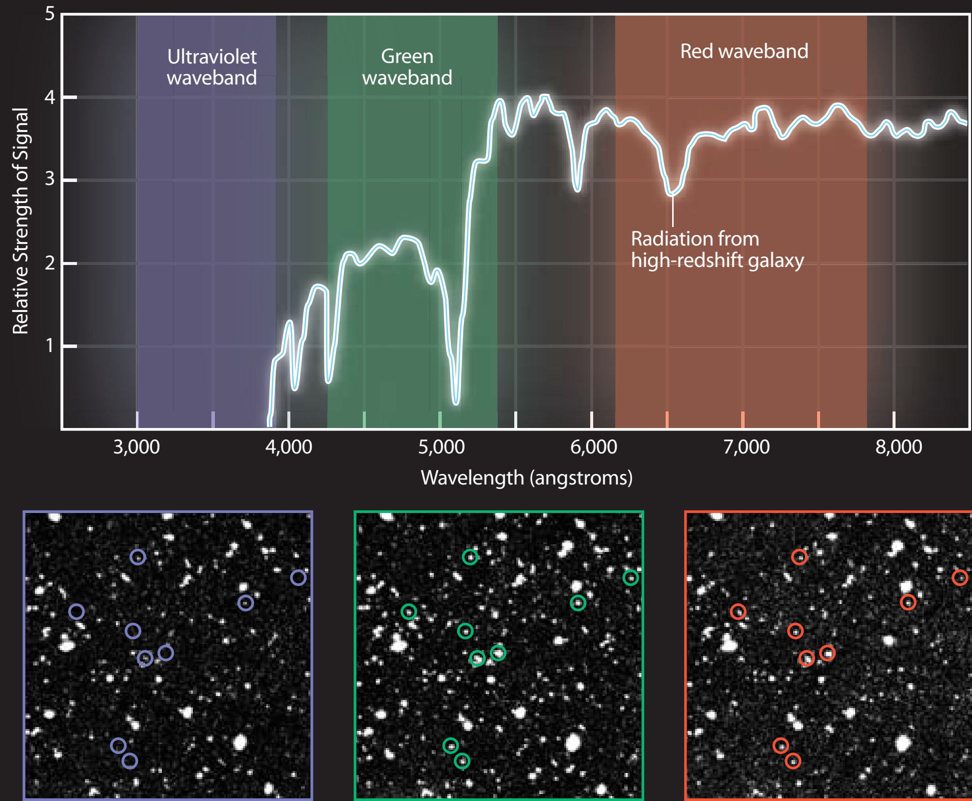
Until recently, astronomers found it difficult to make submillimeter observations with ground-based telescopes, partly because water vapor in the atmosphere absorbs signals of that wavelength. But those difficulties were eased with the introduction of the Submillimeter Common-User Bolometer Array (SCUBA), a camera that was installed on the James Clerk Maxwell Telescope on Mauna Kea in 1997. (Located at a height of four kilometers above sea level, the observatory is above 97 percent of the water in the atmosphere.) Several teams of researchers, one of which I led, used SCUBA to directly image regions of the sky with sufficient sensitivity and area coverage to discover distant, exceptionally luminous dust-obscured sources. Because the resolution is fairly coarse, the galaxies have a bloblike appearance [*see illustration above*]. They are also relatively rare—even after many hours of exposure, few sources appeared on each SCUBA image—but they are among the most luminous galaxies in the universe. It is sobering to realize that before SCUBA became available, we did not even know that these powerful, distant systems existed! Their star formation rates are hundreds of times greater than those of present-day galaxies, another indication that the universe used to be much more exciting than it is now.

New observations make it clear that reports of the **universe's demise** have been greatly exaggerated.

Finding all this previously hidden star formation was revolutionary, but might the universe be covering up other violent activity? For example, gas and dust within galaxies could also be obscuring the radiation emitted by the disks of material whirling around supermassive black holes (those weighing as much as billions of suns). These

Finding Ancient Galaxies

To efficiently detect the oldest galaxies in a survey field, astronomers have developed a technique employing filters that select wavebands in the ultraviolet, green and red parts of the spectrum. Because of the expansion of the universe, the light from the oldest galaxies has been shifted toward the red end; the graph (*top*) shows how a relatively high redshift (about three) can push the radiation from a distant galaxy out of the ultraviolet waveband. As a result, the ancient galaxies appear in images made with the red and green filters but not in images made with the ultraviolet filter (*bottom*).



disks are believed to be the power sources of quasars, the prodigiously luminous objects found at high redshifts, as well as the active nuclei at the centers of many nearby galaxies. Optical studies in the 1980s suggested that there were far more quasars several billion years after the big bang than there are active galactic nuclei in the present-day universe. Because the supermassive black holes that powered the distant quasar activity cannot be destroyed, astronomers presumed that many nearby galaxies must contain dead quasars—black holes that have exhausted their fuel supply.

These dormant supermassive black holes have indeed been detected through their gravitational influence. Stars and gas continue to orbit around the holes even though little material is swirling into them. In fact, a nearly dormant black hole resides at the center of the Milky Way. Together these results led scientists to develop a scenario: most supermassive black holes formed during the quasar era, consumed all the material surrounding them in a violent fit of growth and then disappeared from optical observations once their fuel supply ran out. In short, quasar activity, like star formation, was more vigorous in the distant past, a third sign that we live in relatively boring times.

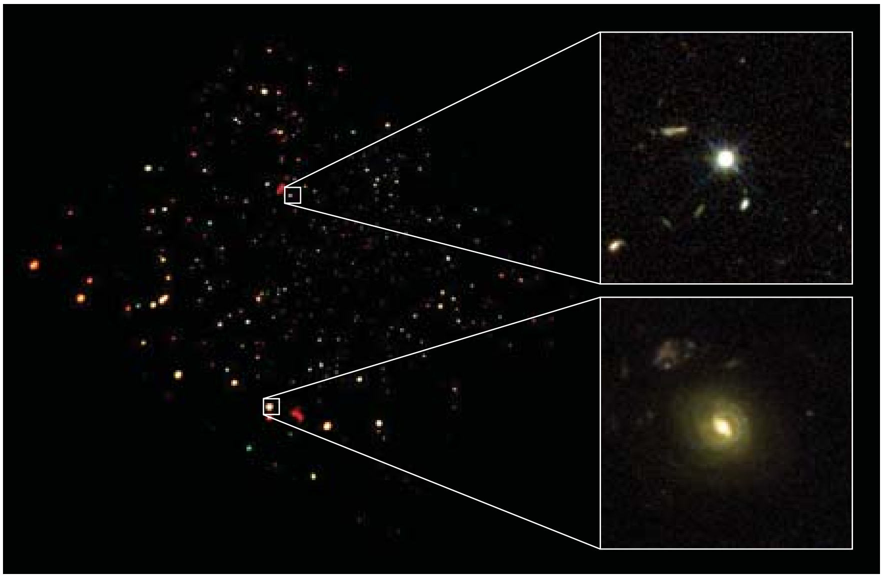
This scenario, however, is incomplete. By combining x-ray and visible-light observations, astronomers are now revisiting the conclusion that the vast majority of quasars died out long ago. X-rays are important because, unlike visible light, they can pass through the gas and dust surrounding hidden black holes. But x-rays are blocked by the earth's atmosphere, so researchers must rely on space telescopes such as the Chandra and XMM-Newton X-ray observatories to detect black hole activity. In 2000 a team consisting of Cowie, Richard F. Mushotzky of the NASA Goddard Space Flight Center, Eric A. Richards, then at Arizona State University, and I used the Subaru telescope at Mauna Kea to identify optical counterparts to 20 x-ray sources found by Chandra in a survey field. We then employed the 10-meter Keck telescope to obtain the spectra of these objects.

Our result was quite unexpected: many of the active supermassive black holes detected by Chandra reside in relatively nearby, luminous galaxies. Modelers of the cosmic x-ray background had predicted the existence of a large population of obscured supermassive black holes, but they had not expected

THE AUTHOR

AMY J. BARGER studies the evolution of the universe by observing some of its oldest objects. She is an associate professor of astronomy at the University of Wisconsin–Madison and also holds an affiliate graduate faculty appointment at the University of Hawaii at Manoa. Barger earned her Ph.D. in astronomy in 1997 at the University of Cambridge, then did postdoctoral research at the University of Hawaii Institute for Astronomy. An observational cosmologist, she has explored the high-redshift universe using the Chandra X-ray Observatory, the Hubble Space Telescope, and the telescopes on Kitt Peak in Arizona and on Mauna Kea in Hawaii.

TOMMY MOORMAN (*graph*); COURTESY OF CHARLES C. STEIDEL *California Institute of Technology* (*images*)



X-RAY VISION can be used to find hidden black holes. The Chandra X-ray Observatory detected many black holes in its Deep Field North survey (*left*). Some were ancient, powering brilliant quasars that flourished just a few billion years after the big bang (*top right*). But others lurked in the centers of relatively nearby galaxies, still generating x-rays in the modern era (*bottom right*).

them to be so close at hand! Moreover, the optical spectra of many of these galaxies showed absolutely no evidence of black hole activity; without the x-ray observations, astronomers could never have discovered the supermassive black holes lurking in their cores.

This research suggests that not all supermassive black holes were formed in the quasar era. These mighty objects have apparently been assembling from the earliest times until the present. The supermassive black holes that are still active, however, do not exhibit the same behavioral patterns as the distant quasars. Quasars are voracious consumers, greedily gobbling up the material around them at an enormous rate. In contrast, most of the nearby sources that Chandra detected are more moderate eaters and thus radiate less intensely. Scientists have not yet determined what mechanism is responsible for this vastly different behavior. One possibility is that the present-day black holes have less gas to consume. Nearby galaxies undergo fewer collisions than the distant, ancient galaxies did, and such collisions could drive material into the supermassive black holes at the galactic centers.

Chandra had yet another secret to reveal: although the moderate x-ray sources were much less luminous than the quasars—generating as little as 1 percent of the radiation emitted by their older counterparts—when we added up the light produced by all the moderate sources in recent times, we found the amount to be about one tenth of that produced by the quasars in early times. The only way this result could arise is if there are many more moderate black holes active now than there were quasars active in the past. In other words, the contents of the universe have transitioned from a small number of bright objects to a large number of dimmer ones. Even though supermassive black holes are now being built smaller and cheaper, their combined effect is still potent.

Star-forming galaxies have also undergone a cosmic downsizing. Although some nearby galaxies are just as extravagant in their star-forming habits as the extremely luminous, dust-obscured galaxies found in the SCUBA images, the density of ultraluminous galaxies in the present-day universe is more than 400 times lower than their density in the distant universe. Again, however, smaller galaxies have taken up some of the slack. A team consisting of Cowie, Gillian Wilson, now at the California Institute of Technology's Spitzer Science Center, Doug J. Burke, now at the Harvard-Smithsonian Center for Astrophysics, and I has refined the estimates of the universe's luminosity density by studying high-quality images produced with a wide range of filters and performing a complete spectroscopic follow-up. We found that the luminosity density of optical and ultraviolet light has not changed all that much with cosmic time. Although the overall star formation rate has dropped in the second half of the universe's lifetime because the monstrous dusty galaxies are no longer bursting with stars, the population of small, nearby star-forming galaxies is so numerous that the density of optical and ultraviolet light is declining rather gradually. This result gives us a much more optimistic outlook on the continuing health of the universe.

Middle-Aged Vigor

THE EMERGING PICTURE of continued vigor fits well with cosmological theory. New computer simulations suggest that the shift from a universe dominated by a few large and powerful galaxies to a universe filled with many smaller and meeker galaxies may be a direct consequence of cosmic expansion. As the universe expands, galaxies become more separated and mergers become rarer. Furthermore, as the gas surrounding galaxies grows more diffuse, it becomes easier to heat. Because hot gas is more energetic than cold gas, it does not gravitationally collapse as readily into the galaxy's potential well. Fabrizio Nicastro of the Harvard-Smithsonian Center for Astrophysics and his co-workers have recently detected a warm intergalactic fog through its absorption of ultraviolet light and x-rays from distant quasars and active galactic nuclei. This warm fog surrounds our galaxy in every direction and is part of the Local Group of galax-

D. M. ALEXANDER, F. E. BAUER, W. N. BRANDT, G. P. GARMIRE ET AL. *NASA/Pennsylvania State University (Chandra image)*; A. J. BARGER AND M. GIAVALISCO *NASA/Great Observatories Origins Deep Survey (optical images)* ies, which includes the Milky Way, Andromeda and 30 smaller galaxies. Most likely this gaseous material was left over from the galaxy formation process but is too warm to permit further galaxy formation to take place.

Small galaxies may lie in cooler environments because they may not have heated their surrounding regions of gas to the same extent that the big galaxies did through supernova explosions and quasar energy. Also, the small galaxies may have consumed less of their surrounding material, allowing them to continue their more modest lifestyles to the present day. In contrast, the larger and more profligate galaxies have exhausted their resources and are no longer able to collect more from their environments. Ongoing observational studies of the gaseous properties of small, nearby galaxies may reveal how they interact with their environments and thus provide a key to understanding galactic evolution.

But a crucial part of the puzzle remains unsolved: How did the universe form monster quasars so early in its history? The Sloan Digital Sky Survey, a major astronomical project to map one quarter of the entire sky and measure distances to more than a million remote objects, has discovered quasars that existed when the universe was only one sixteenth of its present age, about 800 million years after the big bang. In 2003 Fabian Walter, then at the National Radio Astronomy Observatory, and his collaborators detected the presence of carbon monoxide in the emission from one of these quasars; because carbon and oxygen could have been created only from the thermonuclear reactions in stars, this discovery suggests that a significant amount of star formation occurred in the universe's first several hundred million years. Recent results from the Wilkinson Microwave Anisotropy Probe, a satellite that studies the cosmic background radiation, also indicate that star formation began just 400 million years after the big bang.

Furthermore, computer simulations have shown that the first stars were most likely hundreds of times as massive as the sun. Such stars would have burned so brightly that they would have run out of fuel in just a few tens of millions of years; then the heaviest stars would have collapsed to black holes, which could have formed the seeds of the supermassive black holes that powered the first quasars. This explanation for the early appearance of quasars may be bolstered by the further study of gamma-ray bursts, which are believed to result from the collapse of very massive stars into black holes. Because gamma-ray bursts are the most powerful explosions in the universe since the big bang, astronomers can detect them at very great distances. In November 2004 NASA launched the Swift Gamma-Ray Burst Mission, a $250-million satellite with three telescopes designed to observe the explosions and their afterglows in the gamma-ray, x-ray, ultraviolet and optical wavelengths. In the two years since its launch, Swift has identified a number of gamma-ray bursts. The most exciting of these was the September 2005 discovery of an explosion that took place only 900 million years after the big bang. The hope is that this is just the first of many such detections stretching to even greater distances, thereby providing scientists with a much better understanding of how collapsing stars could have started the growth of supermassive black holes in the early universe.

In comic books, Superman looked through walls with his x-ray vision. Astronomers have now acquired a similar ability with the Chandra and XMM-Newton observatories and are making good use of it to peer deep into the dust-enshrouded regions of the universe. What is being revealed is a dramatic transition from the mighty to the meek. The giant star-forming galaxies and voracious black holes of the universe's past are now moribund. A few billion years from now, the smaller galaxies that are active today will have consumed much of their fuel, and the total cosmic output of radiation will decline dramatically. Even our own Milky Way will someday face this same fate. As the cosmic downsizing continues, the dwarf galaxies—which hold only a few million stars each but are the most numerous type of galaxy in the universe—will become the primary hot spots of star formation. Inevitably, though, the universe will darken, and its only contents will be the fossils of galaxies from its glorious past. Old galaxies never die, they just fade away. SA

What is being revealed is a dramatic transition from the mighty to the meek. Dwarf galaxies will become the **primary hot spots** of star formation.

MORE TO EXPLORE

Star Formation History since z = 1.5 as Inferred from Rest-Frame Ultraviolet Luminosity Density Evolution. Gillian Wilson et al. in *Astronomical Journal,* Vol. 124, pages 1258–1265; September 2002. Available online at **arxiv.org/abs/astro-ph/0203168**

Supermassive Black Holes in the Distant Universe. Edited by Amy J. Barger. Astrophysics and Space Science Library, Vol. 308. Springer, 2004.

The Cosmic Evolution of Hard X-ray Selected Active Galactic Nuclei. Amy J. Barger et al. in *Astronomical Journal,* Vol. 129, pages 578–609; February 2005. **arxiv.org/abs/astro-ph/0410527**

information in the HOLOGRAPHIC UNIVERSE

Theoretical results about black holes suggest that the universe could be like a gigantic hologram

By Jacob D. Bekenstein

Illustrations by Alfred T. Kamajian

Ask anybody what the physical world is made of, and you are likely to be told "matter and energy." Yet if we have learned anything from engineering, biology and physics, information is just as crucial an ingredient. The robot at the automobile factory is supplied with metal and plastic but can make nothing useful without copious instructions telling it which part to weld to what and so on. A ribosome in a cell in your body is supplied with amino acid building blocks and is powered by energy released by the conversion of ATP to ADP, but it can synthesize no proteins without the information brought to it from the DNA in the cell's nucleus. Likewise, a century of developments in physics has taught us that information is a crucial player in physical systems and processes. Indeed, a current trend, initiated by John A. Wheeler of Princeton University, is to regard the physical world as made of information, with energy and matter as incidentals.

This viewpoint invites a new look at venerable questions. The information storage capacity of devices such as hard-disk drives has been increasing by leaps and bounds. When will such progress halt? What is the ultimate information capacity of a device that weighs, say, less than a gram and can fit inside a cubic centimeter (roughly the size of a computer chip)? How much information does it take to describe a whole universe? Could that description fit in a computer's memory? Could we, as William Blake memorably penned, "see a world in a grain of sand," or is that idea no more than poetic license?

Remarkably, recent developments in theoretical physics answer some of these questions, and the answers might be important clues to the ultimate theory of reality. By studying the mysterious properties of black holes, physicists have deduced absolute limits on how much information a region of space or a quantity of matter and energy can hold. Related results suggest that our universe, which we perceive to have three spatial dimensions, might instead be "written" on a two-dimensional surface, like a hologram. Our everyday perceptions of the world as three-dimensional would then be either a profound illusion or merely one of two alternative ways of viewing reality. A grain of sand may not encompass our world, but a flat screen might.

A Tale of Two Entropies

FORMAL INFORMATION theory originated in seminal 1948 papers by American applied mathematician Claude E. Shannon, who introduced today's most widely used measure of information content: entropy. Entropy had long been a central concept of thermodynamics, the branch of physics dealing with heat. Thermodynamic entropy is popularly described as the disorder in a physical system. In 1877 Austrian physicist Ludwig Boltzmann characterized it more precisely in terms of the number of distinct microscopic states that the particles composing a chunk of matter could be in while still looking like the same macroscopic chunk of matter. For example, for the air in the room around you, one would count all the ways that the individual gas molecules could be distributed in the room and all the ways they could be moving.

When Shannon cast about for a way to quantify the information contained in, say, a message, he was led by logic to a formula with the same form as Boltzmann's. The Shannon entropy of a message is the number of binary digits, or bits, needed to encode it. Shannon entropy does not enlighten us about the value of information, which is highly dependent on context. Yet as an objective measure of quantity of information, it has been enormously useful in science and technology. For instance, the design of every modern communications device—from cellular phones to modems to compact-disc players—relies on Shannon entropy.

Thermodynamic entropy and Shannon entropy are conceptually equivalent: the number of arrangements that are counted by Boltzmann entropy reflects the amount of Shannon information one would need to implement any particular arrangement. The two entropies have two salient differences, though. First, the thermodynamic entropy used by a chemist or a refrigeration engineer is expressed in units of energy divided by temperature, whereas the Shannon entropy used by a communications engineer is in bits, essentially dimensionless. That difference is merely a matter of convention.

Even when reduced to common units, however, typical values of the two entropies differ vastly in magnitude. A silicon microchip carrying a gigabyte of data, for instance, has a Shannon entropy of about 10^{10} bits (one byte is eight bits), tremendously smaller than the chip's thermodynamic entropy, which is about 10^{23} bits at room temperature. This discrepancy occurs because the entropies are computed for different degrees of freedom. A degree of freedom is any quantity that can vary, such as a coordinate specifying a particle's location or one component of its velocity. The Shannon entropy of the chip cares only about the overall state of each tiny transistor etched in the silicon crystal—the transistor is on or off; it is a 0 or a 1—a single binary degree of freedom.

Thermodynamic entropy, in contrast, depends on the states of all the billions of atoms (and their roaming electrons) that make up each transistor. As miniaturization brings closer the day when each atom will store one bit of information for us, the useful Shannon entropy of the state-of-the-art microchip will edge closer in magnitude to its material's thermodynamic entropy. When the two entropies are calculated for the same degrees of freedom, they are equal.

What are the ultimate degrees of freedom? Atoms, after all, are made of electrons and nuclei, nuclei are agglomerations of protons and neutrons, and those in turn are composed of quarks. Many physicists today consider electrons and quarks to be excitations of superstrings, which they hypothesize to be the most fundamental entities. But the vicissitudes of a century of revelations in physics warn us not to be dogmatic. There could be more levels of structure in our universe than are dreamt of in today's physics.

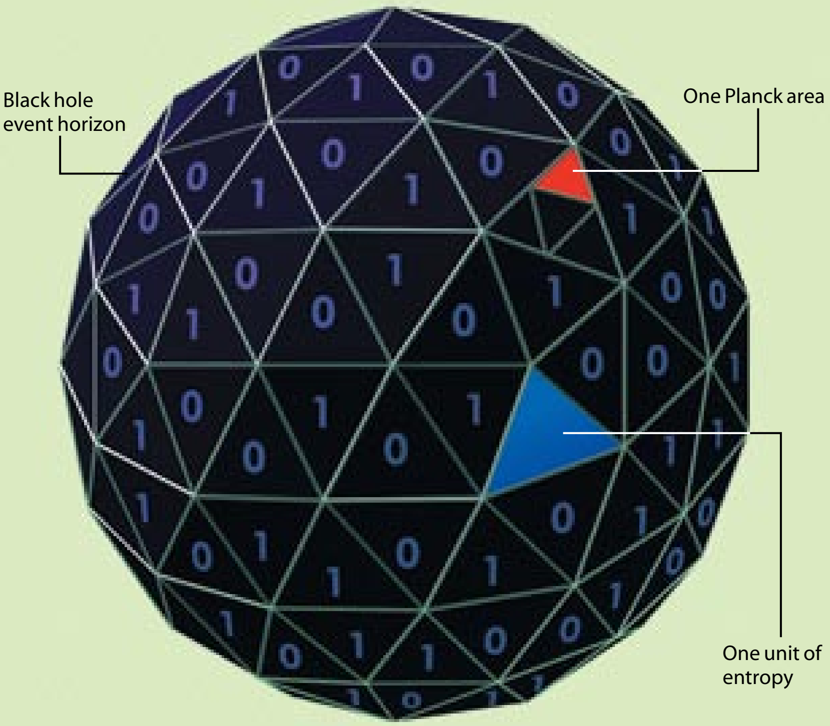
One cannot calculate the ultimate information capacity of a chunk of matter or, equivalently, its true thermodynamic entropy, without knowing the nature of the ultimate constituents of matter or of the deepest level of structure, which I shall refer to as level X. (This ambiguity causes no problems in analyzing practical thermodynamics, such as that of car engines, for example, because the quarks within the atoms can be ignored—they do not change their states under the relatively benign conditions in the engine.) Given the dizzying progress in miniaturization, one can playfully contemplate a day when quarks will serve to store information, one bit apiece perhaps. How much information would then fit into our one-centimeter cube? And how much if we harness superstrings or even deeper, yet undreamt of levels? Surprisingly, developments in gravitation physics in the past three decades have supplied some clear answers to what seem to be elusive questions.

Black Hole Thermodynamics

A CENTRAL PLAYER in these developments is the black hole. Black holes are a consequence of general relativity, Albert Einstein's geometric theory of gravitation, initially published in 1915.

Overview
The World as a Hologram

- An astonishing theory called the holographic principle holds that the universe is like a hologram: just as a trick of light allows a fully three-dimensional image to be recorded on a flat piece of film, our seemingly three-dimensional universe could be completely equivalent to alternative quantum fields and physical laws "painted" on a distant, vast surface.
- The physics of black holes—immensely dense concentrations of mass—provides a hint that the principle might be true. Studies of black holes show that, although it defies common sense, the maximum entropy or information content of any region of space is defined not by its volume but by its surface area.
- Physicists hope that this surprising finding is a clue to the ultimate theory of reality.



ENTROPY OF A BLACK HOLE is proportional to the area of its event horizon, the surface from within which even light cannot escape the gravity of the hole. Specifically, a hole with a horizon spanning A Planck areas has $A/4$ units of entropy. (The Planck area, approximately 10^{-66} square centimeter, is the fundamental quantum unit of area determined by the strength of gravity, the speed of light and the size of quanta.) Considered as information, it is as if the entropy were written on the event horizon, with each bit (each digital 1 or 0) corresponding to four Planck areas.

In this theory, gravitation arises from the curvature of spacetime, which makes objects move as if they were pulled by a force. Conversely, the curvature is caused by the presence of matter and energy. According to Einstein's equations, a sufficiently dense concentration of matter or energy will curve spacetime so extremely that it rends, forming a black hole. The laws of relativity forbid anything that went into a black hole from coming out again, at least within the classical (nonquantum) description of the physics. The point of no return, called the event horizon of the black hole, is of crucial importance. In the simplest case, the horizon is a sphere, whose surface area is larger for more massive black holes.

It is impossible to determine what is inside a black hole. No detailed information can emerge across the horizon and escape into the outside world. In disappearing forever into a black hole, however, a piece of matter does leave some traces. Its energy (we count any mass as energy in accordance with Einstein's $E = mc^2$) is permanently reflected in an increment in the black hole's mass. If the matter is captured while circling the hole, its associated angular momentum is added to the black hole's angular momentum. Both the mass and angular momentum of a black hole are measurable from their effects on spacetime around the hole. In this way, the laws of conservation of energy and angular momentum are upheld by black holes. Another fundamental law, the second law of thermodynamics, appears to be violated.

The second law of thermodynamics summarizes the familiar observation that most processes in nature are irreversible: a teacup falls from the table and shatters, but no one has ever seen shards jump up of their own accord and assemble into a teacup. The second law of thermodynamics forbids such inverse processes. It states that the entropy of an isolated physical system can never decrease; at best, entropy remains constant, and usually it increases. This law is central to physical chemistry and engineering; it is arguably the physical law with the greatest impact outside physics.

As first emphasized by Wheeler, when matter disappears into a black hole, its entropy is gone for good, and the second law seems to be transcended, made irrelevant. A clue to resolving this puzzle came in 1970, when Demetrious Christodoulou, then a graduate student of Wheeler's at Princeton, and Stephen Hawking of the University of Cambridge independently proved that in various processes, such as black hole mergers, the total area of the event horizons never decreases. The analogy with the tendency of entropy to increase led me to propose in 1972 that a black hole has entropy proportional to the area of its horizon [*see illustration on this page*]. I conjectured that when matter falls into a black hole, the increase in black hole entropy always compensates or overcompensates for the "lost" entropy of the matter. More generally, the sum of black hole entropies and the ordinary entropy outside the black holes cannot decrease. This is the generalized second law—GSL for short.

The GSL has passed a large number of stringent, if purely theoretical, tests. When a star collapses to form a black hole, the black hole entropy greatly exceeds the star's entropy. In 1974 Hawking demonstrated that a black hole spontaneously emits thermal radiation, now known as Hawking radiation, by a quantum process. The Christodoulou-

THE AUTHOR

JACOB D. BEKENSTEIN has contributed to the foundation of black hole thermodynamics and to other aspects of the connections between information and gravitation. He is Polak Professor of Theoretical Physics at the Hebrew University of Jerusalem, a member of the Israel Academy of Sciences and Humanities, and a recipient of the Rothschild and the Israel prizes. Bekenstein dedicates this article to John Archibald Wheeler (his Ph.D. supervisor 30 years ago). Wheeler belongs to the third generation of Ludwig Boltzmann's students: Wheeler's Ph.D. adviser, Karl Herzfeld, was a student of Boltzmann's student Friedrich Hasenöhrl.

Limits on Information Density

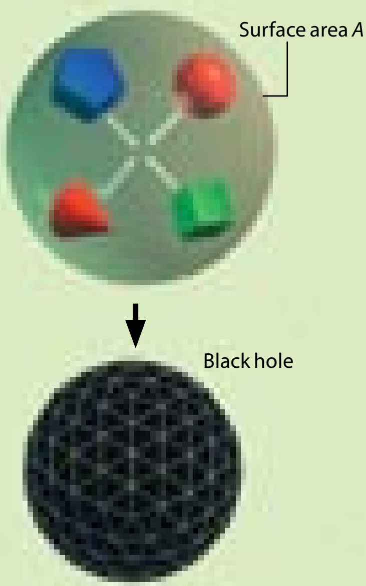
The thermodynamics of black holes allows one to deduce limits on the density of entropy or information in various circumstances.

The holographic bound defines how much information can be contained in a specified region of space. It can be derived by considering a roughly spherical distribution of matter that is contained within a surface of area A. The matter is induced to collapse to form a black hole (*a*). The black hole's area must be smaller than A, so its entropy must be less than $A/4$ [*see illustration on preceding page*]. Because entropy cannot decrease, one infers that the original distribution of matter also must carry less than $A/4$ units of entropy or information. This result—that the maximum information content of a region of space is fixed by its area—defies the commonsense expectation that the capacity of a region should depend on its volume.

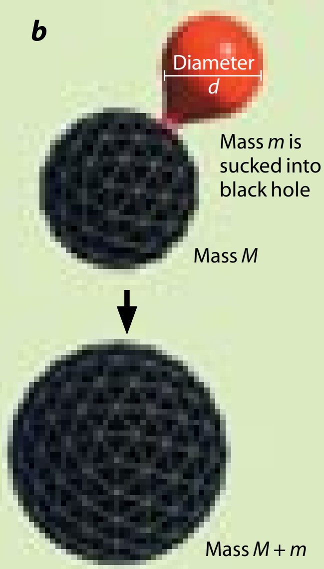
The universal entropy bound defines how much information can be carried by a mass m of diameter d. It is derived by imagining that a capsule of matter is engulfed by a black hole not much wider than it (*b*). The increase in the black hole's size places a limit on how much entropy the capsule could have contained. This limit is tighter than the holographic bound, except when the capsule is almost as dense as a black hole (in which case the two bounds are equivalent).

The holographic and universal information bounds are far beyond the data storage capacities of any current technology, and they greatly exceed the density of information on chromosomes and the thermodynamic entropy of water (*c*).—*J.D.B.*

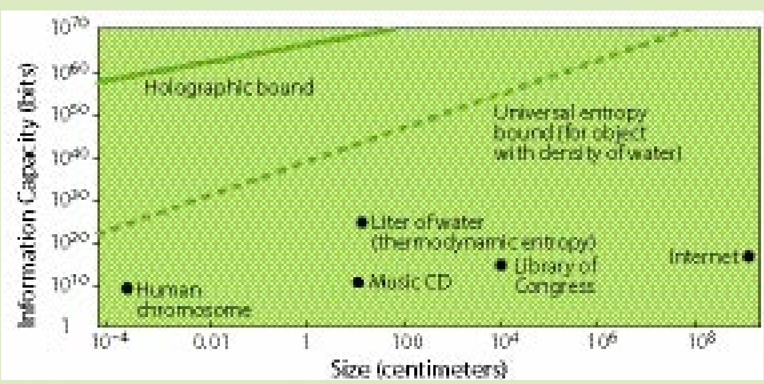
a



b



c



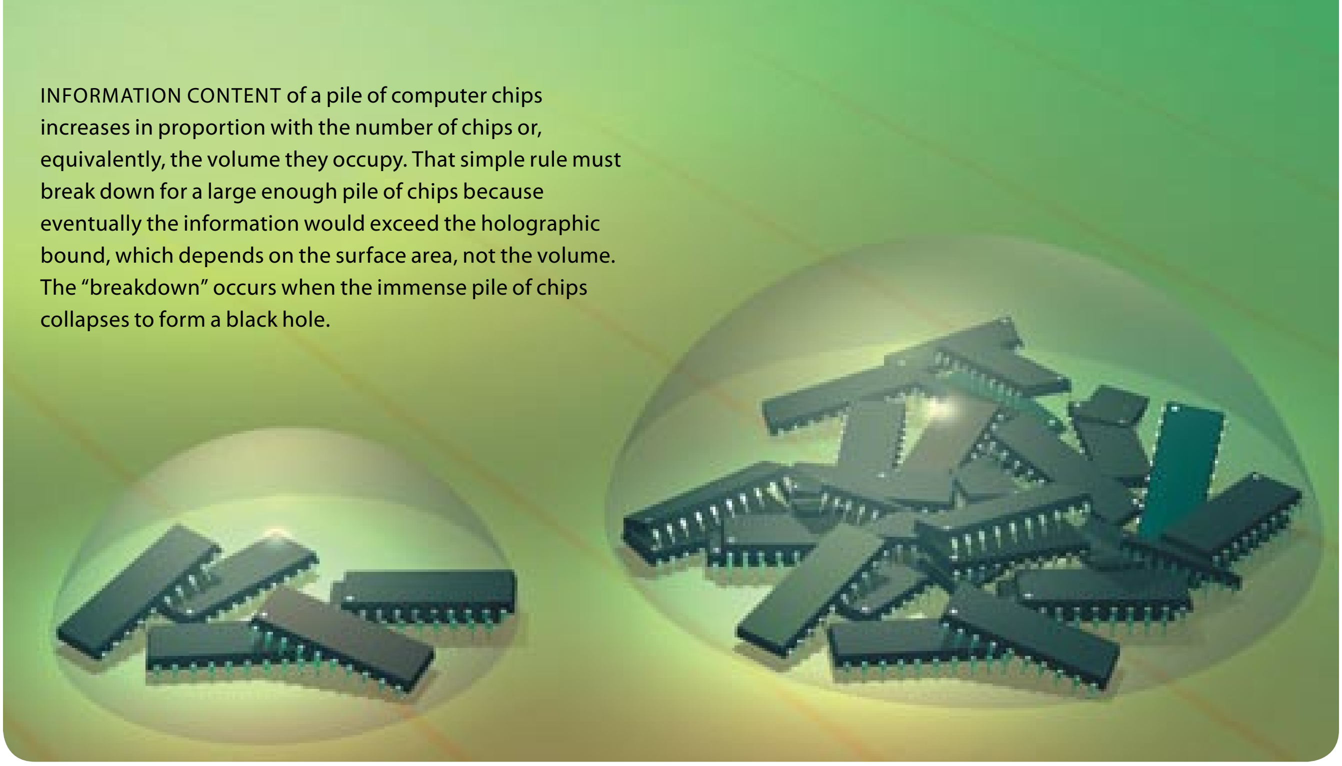
Hawking theorem fails in the face of this phenomenon (the mass of the black hole, and therefore its horizon area, decreases), but the GSL copes with it: the entropy of the emergent radiation more than compensates for the decrement in black hole entropy, so the GSL is preserved. In 1986 Rafael D. Sorkin of Syracuse University exploited the horizon's role in barring information inside the black hole from influencing affairs outside to show that the GSL (or something very similar to it) must be valid for any conceivable process that black holes undergo. His deep argument makes it clear that the entropy entering the GSL is that calculated down to level X, whatever that level may be.

Hawking's radiation process allowed him to determine the proportionality constant between black hole entropy and horizon area: black hole entropy is precisely one quarter of the event horizon's area measured in Planck areas. (The Planck length, about 10^{-33} centimeter, is the fundamental length scale related to gravity and quantum mechanics. The Planck area is its square.) Even in thermodynamic terms, this is a vast quantity of entropy. The entropy of a black hole one centimeter in diameter would be about 10^{66} bits, roughly equal to the thermodynamic entropy of a cube of water 10 billion kilometers on a side.

The World as a Hologram

THE GSL ALLOWS US to set bounds on the information capacity of any isolated physical system, limits that refer to the information at all levels of structure down to level X. In 1980 I began studying the first such bound, called the universal entropy bound, which limits how much entropy can be carried by a specified mass of a specified size [*see box at left*]. A related idea, the holographic bound, was foreshadowed in 1993 by Nobel laureate Gerard 't Hooft of the University of Utrecht in the Netherlands and developed in 1995 by Leonard Susskind of Stanford University. It limits how much entropy can be contained in matter and energy occupying a specified volume of space.

LAURIE GRACE (*graph*)



INFORMATION CONTENT of a pile of computer chips increases in proportion with the number of chips or, equivalently, the volume they occupy. That simple rule must break down for a large enough pile of chips because eventually the information would exceed the holographic bound, which depends on the surface area, not the volume. The "breakdown" occurs when the immense pile of chips collapses to form a black hole.

In his work on the holographic bound, Susskind considered any approximately spherical isolated mass that is not itself a black hole and that fits inside a closed surface of area A. If the mass can collapse to a black hole, that hole will end up with a horizon area smaller than A. The black hole entropy is therefore smaller than $A/4$. According to the GSL, the entropy of the system cannot decrease, so the mass's original entropy cannot have been bigger than $A/4$. It follows that the entropy of an isolated physical system with boundary area A is necessarily less than $A/4$. What if the mass does not spontaneously collapse? In 2000 I showed that a tiny black hole can be used to convert the system to a black hole not much different from the one in Susskind's argument. The bound is therefore independent of the constitution of the system or of the nature of level X. It just depends on the GSL.

We can now answer some of those elusive questions about the ultimate limits of information storage. A device measuring a centimeter across could in principle hold up to 10^{66} bits—a mind-boggling amount. The visible universe contains at least 10^{100} bits of entropy, which could in principle be packed inside a sphere a tenth of a light-year across. Estimating the entropy of the universe is a difficult problem, however, and much larger numbers, requiring a sphere almost as big as the universe itself, are entirely plausible.

But it is another aspect of the holographic bound that is truly astonishing. Namely, that the maximum possible entropy depends on the boundary area instead of the volume. Imagine that we are piling up computer memory chips in a big heap. The number of transistors—the total data storage capacity—increases with the volume of the heap. So, too, does the total thermodynamic entropy of all the chips. Remarkably, though, the theoretical ultimate information capacity of the space occupied by the heap increases only with the surface area. Because volume increases more rapidly than surface area, at some point the entropy of all the chips would exceed the holographic bound. It would seem that either the GSL or our commonsense ideas of entropy and information capacity must fail. In fact, what fails is the pile itself: it would collapse under its own gravity and form a black hole before that impasse was reached. Thereafter each additional memory chip would increase the mass and surface area of the black hole in a way that would continue to preserve the GSL.

This surprising result—that information capacity depends on surface area—has a natural explanation if the holographic *principle* (proposed by 't Hooft and elaborated by Susskind) is true. In the everyday world, a hologram is a special kind of photograph that generates a full three-dimensional image when it is illuminated in the right manner. All the information describing the 3-D scene is encoded into the pattern of light and dark areas on the two-dimensional piece of film, ready to be regenerated. The holographic principle contends that an analogue of this visual magic applies to the full physical description of any system occupying a 3-D region: it proposes that another physical theory defined only on the 2-D boundary of the region completely describes the 3-D physics. If a 3-D system can be fully described by a physical theory operating solely on its 2-D boundary, one would expect the information content of the system not to exceed that of the description on the boundary.

A Universe Painted on Its Boundary

CAN WE APPLY the holographic principle to the universe at large? The real universe is a 4-D system: it has volume and extends in time. If the physics of our universe is holographic, there would be an alternative set of physical laws, operating on a 3-D boundary of spacetime somewhere, that would be equivalent to our known 4-D physics. We do not yet know of any such 3-D theory that works in that way. Indeed, what surface should we use as the boundary of the universe? One step toward realizing these ideas is to study models that are simpler than our real universe.

A class of concrete examples of the holographic principle at work involves so-called anti–de Sitter spacetimes. The original de Sitter spacetime is a model universe first obtained by Dutch astronomer Willem de Sitter in 1917 as a solution of Einstein's equations, including the repulsive force known as the cosmological constant. De Sitter spacetime is empty, expands at an accelerating rate and is very highly symmetrical. In 1997 astronomers studying distant supernova explosions concluded that our universe now expands in an accelerated fashion and will probably become increasingly like a de Sitter spacetime in the future. Now, if the repulsive cosmological constant is replaced by an attractive one, de Sitter's solution turns into anti–de Sitter spacetime, which has equally as much symmetry. More important for the holographic concept, it possesses a boundary, which is located "at infinity" and is a lot like our everyday spacetime.

Using anti–de Sitter spacetime, theorists have devised a concrete example of the holographic principle at work: a universe described by superstring theory functioning in an anti–de Sitter spacetime is completely equivalent to a quantum field theory operating on the boundary of that spacetime [*see box below*]. Thus, the full majesty of superstring theory in an anti–de Sitter universe is painted on the boundary of the universe. Juan Maldacena, then at Harvard University, first conjectured such a relation in 1997 for the 5-D anti–de Sitter case, and it was later confirmed for many situations by Edward Witten of the Institute for Advanced Study in Princeton, N.J., and Steven S. Gubser, Igor R. Klebanov and Alexander M. Polyakov of Princeton University. Examples of this holographic correspondence are now known for spacetimes with a variety of dimensions.

This result means that two ostensibly very different theories—not even acting in spaces of the same dimension—are equivalent. Creatures living in one of these universes would be incapable of determining if they inhabited a 5-D universe described by string theory or a 4-D one described by a quantum field theory of point particles. (Of course, the structures of their brains might give them an overwhelming "commonsense" prejudice in favor of one description or another, in just the way that our brains construct an innate perception that our universe has three spatial dimensions; see the illustration on the opposite page.)

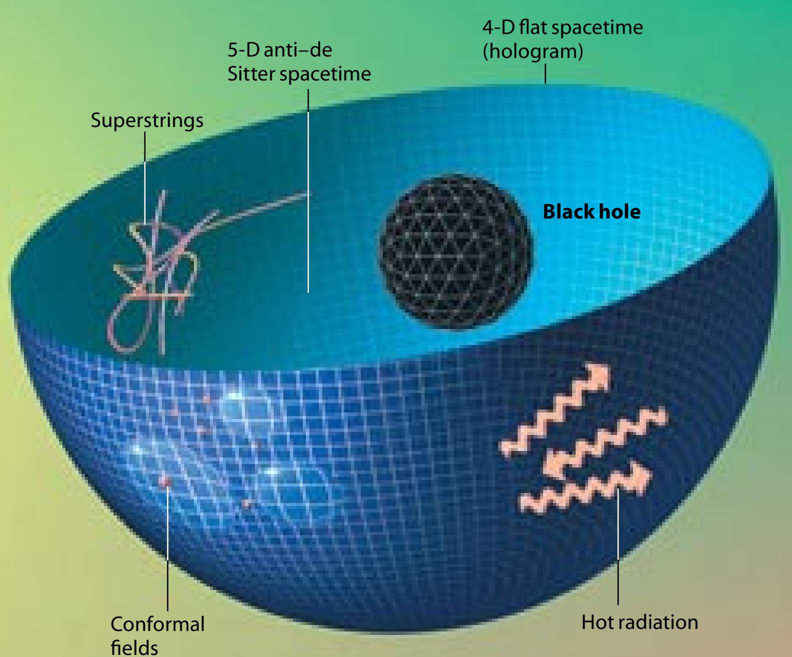
The holographic equivalence can allow a difficult calculation in the 4-D boundary spacetime, such as the behavior of quarks and gluons, to be traded for another, easier calculation in the highly symmetric, 5-D anti–de Sitter spacetime. The correspondence works the other way, too. Witten has shown that a black hole in anti–de Sitter spacetime corresponds to hot radiation in the alternative physics operating on the bounding spacetime. The entropy of the hole—a deeply mysterious concept—equals the radiation's entropy, which is quite mundane.

The Expanding Universe

HIGHLY SYMMETRIC and empty, the 5-D anti–de Sitter universe is hardly like our universe existing in 4-D, filled with matter and radiation and riddled with

A Holographic Spacetime

Two universes of different dimensions and obeying disparate physical laws are rendered completely equivalent by the holographic principle. Theorists have demonstrated this principle mathematically for a specific type of five-dimensional spacetime ("anti–de Sitter") and its four-dimensional boundary. In effect, the 5-D universe is recorded like a hologram on the 4-D surface at its periphery. Superstring theory rules in the 5-D spacetime, but a so-called conformal field theory of point particles operates on the 4-D hologram. A black hole in the 5-D spacetime is equivalent to hot radiation on the hologram—for example, the hole and the radiation have the same entropy even though the physical origin of the entropy is completely different for each case. Although these two descriptions of the universe seem utterly unalike, no experiment could distinguish between them, even in principle. —*J.D.B.*



violent events. Even if we approximate our real universe with one that has matter and radiation spread uniformly throughout, we get not an anti–de Sitter universe but rather a "Friedmann-Robertson-Walker" universe. Most cosmologists today concur that our universe resembles an FRW universe, one that is infinite, has no boundary and will go on expanding ad infinitum.

Does such a universe conform to the holographic principle or the holographic bound? Susskind's argument based on collapse to a black hole is of no help here. Indeed, the holographic bound deduced from black holes must break down in a uniform expanding universe. The entropy of a region uniformly filled with matter and radiation is truly proportional to its volume. A sufficiently large region will therefore violate the holographic bound.

In 1999 Raphael Bousso, then at Stanford, proposed a modified holographic bound, which has since been found to work even in situations where the bounds we discussed earlier cannot be applied. Bousso's formulation starts with any suitable 2-D surface; it may be closed like a sphere or open like a sheet of paper. One then imagines a brief burst of light issuing simultaneously and perpendicularly from all over one side of the surface. The only demand is that the imaginary light rays are converging to start with. Light emitted from the inner surface of a spherical shell, for instance, satisfies that requirement. One then considers the entropy of the matter and radiation that these imaginary rays traverse, up to the points where they start crossing. Bousso conjectured that this entropy cannot exceed the entropy represented by the initial surface—one quarter of its area, measured in Planck areas. This is a different way of tallying up the entropy than that used in the original holographic bound. Bousso's bound refers not to the entropy of a region at one time but rather to the sum of entropies of locales at a variety of times: those that are "illuminated" by the light burst from the surface.

Bousso's bound subsumes other entropy bounds while avoiding their limitations. Both the universal entropy bound and the 't Hooft-Susskind form of the holographic bound can be deduced from Bousso's for any isolated system that is not evolving rapidly and whose gravitational field is not strong. When these conditions are overstepped—as for a collapsing sphere of matter already inside a black hole—these bounds eventually fail, whereas Bousso's bound continues to hold. Bousso has also shown that his strategy can be used to locate the 2-D surfaces on which holograms of the world can be set up.

OUR INNATE PERCEPTION that the world is three-dimensional could be an extraordinary illusion.

Augurs of a Revolution

RESEARCHERS HAVE proposed many other entropy bounds. The proliferation of variations on the holographic motif makes it clear that the subject has not yet reached the status of physical law. But although the holographic way of thinking is not yet fully understood, it seems to be here to stay. And with it comes a realization that the fundamental belief, prevalent for 50 years, that field theory is the ultimate language of physics must give way. Fields, such as the electromagnetic field, vary continuously from point to point, and they thereby describe an infinity of degrees of freedom. Superstring theory also embraces an infinite number of degrees of freedom. Holography restricts the number of degrees of freedom that can be present inside a bounding surface to a finite number; field theory with its infinity cannot be the final story. Furthermore, even if the infinity is tamed, the mysterious dependence of information on surface area must be somehow accommodated.

Holography may be a guide to a better theory. What is the fundamental theory like? The chain of reasoning involving holography suggests to some, notably Lee Smolin of the Perimeter Institute for Theoretical Physics in Waterloo, Ontario, that such a final theory must be concerned not with fields, not even with spacetime, but rather with information exchange among physical processes. If so, the vision of information as the stuff the world is made of will have found a worthy embodiment. SA

MORE TO EXPLORE

Black Hole Thermodynamics. Jacob D. Bekenstein in *Physics Today,* Vol. 33, No. 1, pages 24–31; January 1980.

Black Holes and Time Warps: Einstein's Outrageous Legacy. Kip S. Thorne. W. W. Norton, 1995.

Black Holes and the Information Paradox. Leonard Susskind in *Scientific American,* Vol. 276, No. 4, pages 52–57; April 1997.

Three Roads to Quantum Gravity. Lee Smolin. Basic Books, 2002.

An Introduction to Black Holes, Information and the String Theory Revolution: The Holographic Universe. Leonard Susskind and James Lindesay. World Scientific Publishing, 2005.

HOLOGRAPHIC THEORY relates one set of physical laws acting in a volume with a different set of physical laws acting on a boundary surface, as represented here by the juggler and her colorful two-dimensional image. The surface laws involve quantum particles that have "color" charges and interact very like the quarks and gluons of standard particle physics. The interior laws are a form of string theory and include the force of gravity (experienced by the juggler), which is hard to describe in terms of quantum mechanics. Nevertheless, the physics on the surface and in the interior are completely equivalent, despite their radically different descriptions.

the illusion of GRAVITY

The force of gravity and one of the dimensions of space might be generated out of the peculiar interactions of particles and fields existing in a lower-dimensional realm

By Juan Maldacena

Three spatial dimensions are visible all around us—up/down, left/right, forward/backward. Add time to the mix, and the result is a four-dimensional blending of space and time known as spacetime. Thus, we live in a four-dimensional universe. Or do we?

Amazingly, some new theories of physics predict that one of the three dimensions of space could be a kind of an illusion—that in actuality all the particles and fields that make up reality are moving about in a two-dimensional realm like the Flatland of Edwin A. Abbott. Gravity, too, would be part of the illusion: a force that is not present in the two-dimensional world but that materializes along with the emergence of the illusory third dimension.

Or, more precisely, the theories predict that the number of dimensions in reality could be a matter of perspective: physicists could choose to describe reality as obeying one set of laws (including gravity) in three dimensions or, equivalently, as obeying a different set of laws that operates in two dimensions (in the absence of gravity). Despite the radically different descriptions, both theories would describe everything that we see and all the data we could gather about how the universe works. We would have no way to determine which theory was "really" true.

Such a scenario strains the imagination. Yet an analogous phenomenon occurs in everyday life. A hologram is a two-dimensional object, but when viewed under the correct lighting conditions it produces a fully three-dimensional image. All the information describing the three-dimensional image is in essence encoded in the two-dimensional hologram. Similarly, according to the new physics theories, the entire universe could be a kind of a hologram.

The holographic description is more than just an intellectual or philosophical curiosity. A computation that might be very difficult in one realm can turn out to be relatively straightforward in the other, thereby turning some intractable problems of physics into ones that are easily solved. For example, the theory seems useful in analyzing a recent experimental high-energy physics result. Moreover, the holographic theories offer a fresh way to begin constructing a quantum theory of gravity—a theory of gravity that respects the principles of quantum mechanics. A quantum theory of gravity is a key ingredient in any effort to unify all the forces of nature, and it is needed to explain both what goes on in black holes and what happened in the nanoseconds after the big bang. The holographic theories provide potential resolutions of profound mysteries that have dogged attempts to understand how a theory of quantum gravity could work.

PHOTOGRAPH BY CARY WOLINSKY; PHOTOILLUSTRATION BY BRYAN CHRISTIE DESIGN

A Difficult Marriage

A QUANTUM THEORY of gravity is a holy grail for a certain breed of physicist because all physics except for gravity is well described by quantum laws. The quantum description of physics represents an entire paradigm for physical theories, and it makes no sense for one theory, gravity, to fail to conform to it. Now about 80 years old, quantum mechanics was first developed to describe the behavior of particles and forces in the atomic and subatomic realms. It is at those size scales that quantum effects become significant. In quantum theories, objects do not have definite positions and velocities but instead are described by probabilities and waves that occupy regions of space. In a quantum world, at the most fundamental level everything is in a state of constant flux, even "empty" space, which is in fact filled with virtual particles that perpetually pop in and out of existence.

In contrast, physicists' best theory of gravity, general relativity, is an inherently classical (that is, nonquantum) theory. Albert Einstein's magnum opus, general relativity explains that concentrations of matter or energy cause spacetime to curve and that this curvature deflects the trajectories of particles, just as should happen for particles in a gravitational field. General relativity is a beautiful theory, and many of its predictions have been tested to great accuracy.

In a classical theory such as general relativity, objects have definite locations and velocities, like the planets orbiting the sun. One can plug those locations and velocities (and the masses of the objects) into the equations of general relativity and deduce the curvature of spacetime and from that deduce the effects of gravity on the objects' trajectories. Furthermore, empty spacetime is perfectly smooth no matter how closely one examines it—it is a seamless arena in which matter and energy can play out their lives.

The problem in devising a quantum version of general relativity is not just that on the scale of atoms and electrons, particles do not have definite locations and velocities. To make matters worse, at the even tinier scale delineated by the Planck length (10^{-33} centimeter), quantum principles imply that spacetime itself should be a seething foam, similar to the sea of virtual particles that fills empty space. When matter and spacetime are so protean, what do the equations of general relativity predict? The answer is that the equations are no longer adequate. If we assume that matter obeys the laws of quantum mechanics and gravity obeys the laws of general relativity, we end up with mathematical contradictions. A quantum theory of gravity (one that fits within the paradigm of quantum theories) is needed.

In most situations, the contradictory requirements of quantum mechanics and general relativity do not cause a problem, because either the quantum effects or the gravitational effects are so small that they can be neglected or dealt with by approximations. When the curvature of spacetime is very large, however, the quantum aspects of gravity become significant. It takes a very large mass or a great concentration of mass to produce much spacetime curvature. Even the curvature produced near the sun is exceedingly small compared with the amount needed for quantum gravity effects to become apparent.

Although these effects are completely negligible now, they were very important in the beginning of the big bang, which is why a quantum theory of gravity is needed to describe how the big bang started. Such a theory is also important for understanding what happens at the center of black holes, because matter there is crushed into a region of extremely high curvature. Because gravity involves spacetime curva-

A quantum theory of gravity will probably provide us with an **entirely new perspective** on what spacetime is.

Overview
Equivalent Worlds

- According to a remarkable theory, a universe that exists in two dimensions and is without gravity may be completely equivalent to a three-dimensional universe with gravity. The three-dimensional universe would emerge from the physics of the two-dimensional universe somewhat like a holographic image arising from a hologram.
- The two-dimensional universe exists on the boundary of the three-dimensional universe. The physics on the boundary looks like strongly interacting quarks and gluons. The physics on the interior includes a quantum theory of gravity—something that string theorists have been developing for decades.
- The equivalence provides a new way to understand properties of black holes, which require a suitable melding of quantum mechanics and gravity. The mathematics of the theory has not yet been rigorously proved, but it seems useful in analyzing a recent experimental high-energy physics result.

ture, a quantum gravity theory will also be a theory of quantum spacetime; it should clarify what constitutes the "spacetime foam" mentioned earlier, and it will probably provide us with an entirely new perspective on what spacetime is at the deepest level of reality.

A very promising approach to a quantum theory of gravity is string theory, which some theoretical physicists have been exploring since the 1970s. String theory overcomes some of the obstacles to building a logically consistent quantum theory of gravity. String theory, however, is still under construction and is not yet fully understood. That is, we string theorists have some approximate equations for strings, but we do not know the exact equations. We also do not know the guiding underlying principle that explains the form of the equations, and there are innumerable physical quantities that we do not know how to compute from the equations.

In recent years string theorists have obtained many interesting and surprising results, giving novel ways of understanding what a quantum spacetime is like. I will not describe string theory in much detail here but instead will focus on one of the most exciting developments emerging from string theory research, which has led to a complete, logically consistent, quantum description of gravity in what are called negatively curved spacetimes. For these spacetimes, holographic theories appear to be true.

Negatively Curved Spacetimes

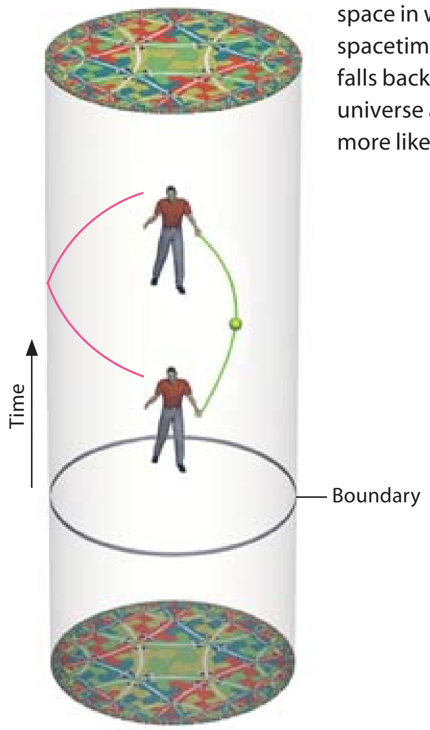
ALL OF US are familiar with Euclidean geometry, where space is flat (that is, not curved). It is the geometry of figures drawn on flat sheets of paper. To a very good approximation, it is also the geometry of the world around us: parallel lines never meet, and all the rest of Euclid's axioms hold.

We are also familiar with some curved spaces. Curvature comes in two forms, positive and negative. The simplest space with positive curvature is the surface of a sphere. A sphere has

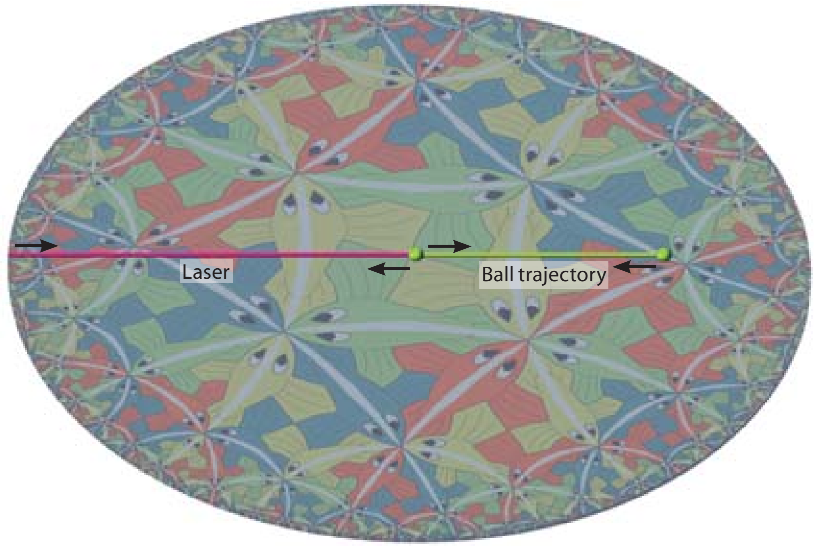
HYPERBOLIC SPACE is depicted in this M. C. Escher drawing (*above*). Each fish is actually the same size, and the circular boundary is infinitely far from the center of the disk. The projection from true hyperbolic space to this representation of it squashes the distant fish to fit the infinite space inside the finite circle. Drawn without that squashing effect, the space is wildly curved, with each small section (*below*) being somewhat like a saddle shape with extra folds.

Negatively Curved Spacetime

The holographic theory involves a negatively curved spacetime known as anti–de Sitter space.



Imagine disks of hyperbolic space stacked atop one another, each representing the state of the universe at one instant. The resulting cylinder is three-dimensional anti–de Sitter space in which the height dimension represents time. Physics operates strangely in such a spacetime: a particle (such as a tennis ball, *green line*) thrown away from the center always falls back in a fixed period, and a laser beam (*red line*) can travel to the boundary of the universe and back in that same interval. In the four-dimensional version, which would be more like our universe, the boundary for each instant would be a sphere instead of a circle.



constant positive curvature. That is, it has the same degree of curvature at every location (unlike an egg, say, which has more curvature at the pointy end).

The simplest space with negative curvature is called hyperbolic space, which is defined as space with constant negative curvature. This kind of space has long fascinated scientists and artists alike. Indeed, M. C. Escher produced several beautiful pictures of hyperbolic space, one of which is shown on the preceding page. His picture is like a flat map of the space. The way that the fish become smaller and smaller is just an artifact of how the curved space is squashed to fit on a flat sheet of paper, similar to the way that countries near the poles get stretched on a map of the globe (a sphere).

By including time in the game, physicists can similarly consider space*times* with positive or negative curvature. The simplest spacetime with positive curvature is called de Sitter space, after Willem de Sitter, the Dutch physicist who introduced it. Many cosmologists believe that the very early universe was close to being a de Sitter space. The far future may also be de Sitter–like because of cosmic acceleration. Conversely, the simplest negatively curved spacetime is called anti–de Sitter space. It is similar to hyperbolic space except that it also contains a time direction. Unlike our universe, which is expanding, anti–de Sitter space is neither expanding nor contracting. It looks the same at all times. Despite that difference, anti–de Sitter space turns out to be quite useful in the quest to form quantum theories of spacetime and gravity.

If we picture hyperbolic space as being a disk like Escher's drawing, then anti–de Sitter space is like a stack of those disks, forming a solid cylinder [*see box above*]. Time runs along the cylinder. Hyperbolic space can have more than two spatial dimensions. The anti–de Sitter space most like our spacetime (with three spatial dimensions) would have a three-dimensional "Escher print" as the cross section of its "cylinder."

Physics in anti–de Sitter space has some strange properties. If you were freely floating anywhere in anti–de Sitter space, you would feel as though you were at the bottom of a gravitational well. Any object that you threw out would come back like a boomerang. Surprisingly, the time required for an object to come back would be independent of how hard you threw it. The difference would just be that the harder you threw it, the farther away it would get on its round-trip back to you. If you sent a flash of light, which consists of

THE AUTHOR

JUAN MALDACENA is a professor in the School of Natural Sciences at the Institute for Advanced Study in Princeton, N.J. From 1997 to 2001 he was in the physics department at Harvard University. He is currently studying various aspects of the duality conjecture described in this article. String theorists were so impressed with the conjecture that at the Strings '98 conference they feted him with a song, *The Maldacena*, sung and danced to the tune of *The Macarena*.

ALFRED T. KAMAJIAN; SILVIO LEVY AND THE UNIVERSITY OF MINNESOTA (*Escher fish*) photons moving at the maximum possible speed (the speed of light), it would actually reach infinity and come back to you, all in a finite amount of time. This can happen because an object experiences a kind of time contraction of ever greater magnitude as it gets farther away from you.

The Hologram

ANTI–DE SITTER SPACE, although it is infinite, has a "boundary," located out at infinity. To draw this boundary, physicists and mathematicians use a distorted length scale similar to Escher's, squeezing an infinite distance into a finite one. This boundary is like the outer circumference of the Escher print or the surface of the solid cylinder I considered earlier. In the cylinder example, the boundary has two dimensions—one is space (looping around the cylinder), and one is time (running along its length). For four-dimensional anti–de Sitter space, the boundary has two space dimensions and one time dimension. Just as the boundary of the Escher print is a circle, the boundary of four-dimensional anti–de Sitter space at any moment in time is a sphere. This boundary is where the hologram of the holographic theory lies.

Stated simply, the idea is as follows: a quantum gravity theory in the interior of an anti–de Sitter spacetime is completely equivalent to an ordinary quantum particle theory living on the boundary. If true, this equivalence means that we can use a quantum particle theory (which is relatively well understood) to define a quantum gravity theory (which is not).

To make an analogy, imagine you have two copies of a movie, one on reels of 70-millimeter film and one on a DVD. The two formats are utterly different, the first a linear ribbon of celluloid with each frame recognizably related to scenes of the movie as we know it, the second a two-dimensional platter with rings of dots that would form a sequence of 0s and 1s if we could perceive them at all. Yet both "describe" the same movie.

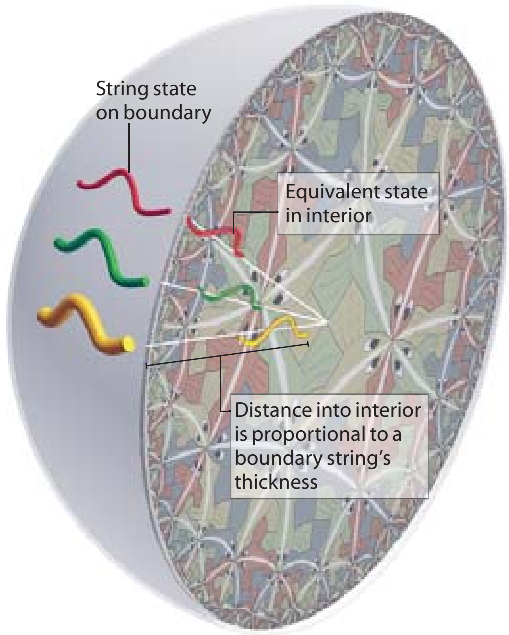
Similarly, the two theories, superficially utterly different in content, describe the same universe. The DVD looks like a metal disk with some glints of rainbowlike patterns. The boundary particle theory "looks like" a theory of particles in the absence of gravity. From the DVD, detailed pictures emerge only when the bits are processed the right way. From the boundary particle theory, quantum gravity and an extra dimension emerge when the equations are analyzed the right way.

What does it really mean for the two theories to be equivalent? First, for every entity in one theory, the other theory has a counterpart. The entities may be very different in how they are described by the theories: one entity in the interior might be a single particle of some type, corresponding on the boundary to a whole collection of particles of another type, considered as one entity. Second, the predictions for corresponding entities must be identical. Thus, if two particles have a 40 percent chance

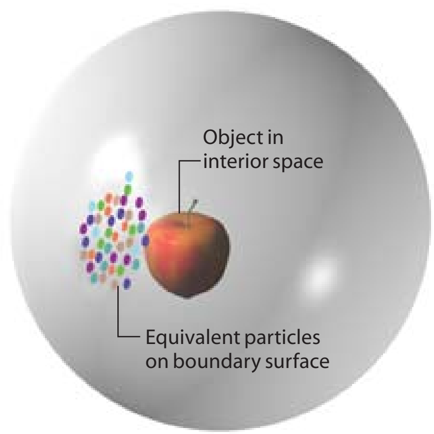
Conjuring a Dimension

Holographic theory describes how quarks and gluons interacting on the boundary of an anti–de Sitter space could be equivalent to particles in the higher-dimensional interior of the space.

Quarks and gluons on the spherical surface of the anti–de Sitter space interact to form strings of various thicknesses. A holographic interpretation of those strings is that in the interior space they represent elementary particles (which are also strings) whose distance from the boundary corresponds to the thicknesses of the strings.



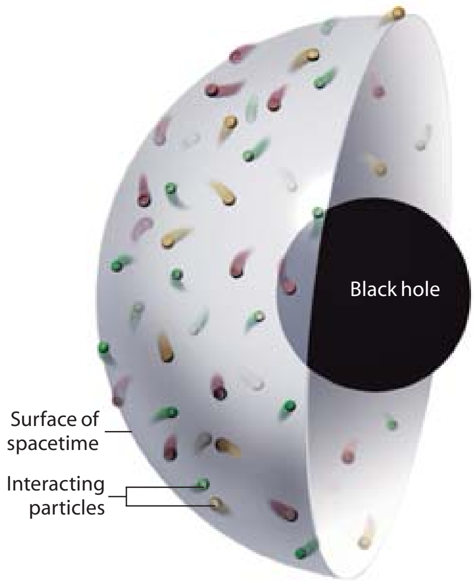
Clouds of quarks and gluons on the boundary surface can thus describe equivalent complex objects (such as this apple) in the interior. The advantage of this holographic theory is that the interior objects experience gravity even though a distinct gravitational interaction does not exist on the surface.



ALFRED T. KAMAJIAN

Understanding Black Holes

Physicist Stephen Hawking showed in the 1970s that black holes have a temperature and give off radiation, but physicists since then have been deeply puzzled. Temperature is a property of a collection of particles, but what is the collection that defines a black hole? The holographic theory solves this puzzle by showing that a black hole is equivalent to a swarm of interacting particles on the boundary surface of spacetime.



of colliding in the interior, the two corresponding collections of particles on the boundary should also have a 40 percent chance of colliding.

Here is the equivalence in more detail. The particles that live on the boundary interact in a way that is very similar to how quarks and gluons interact in reality (quarks are the constituents of protons and neutrons; gluons generate the strong nuclear force that binds the quarks together). Quarks have a kind of charge that comes in three varieties, called colors, and the interaction is called chromodynamics. The difference between the boundary particles and ordinary quarks and gluons is that the particles have a large number of colors, not just three.

Gerard 't Hooft of Utrecht University in the Netherlands studied such theories as long ago as 1974 and predicted that the gluons would form chains that behave much like the strings of string theory. The precise nature of these strings remained elusive, but in 1981 Alexander M. Polyakov, now at Princeton University, noticed that the strings effectively live in a higher-dimensional space than the gluons do. As we shall see shortly, in our holographic theories that higher-dimensional space is the interior of anti–de Sitter space.

To understand where the extra dimension comes from, start by considering one of the gluon strings on the boundary. This string has a thickness, related to how much its gluons are smeared out in space. When physicists calculate how these strings on the boundary of anti–de Sitter space interact with one another, they get a very odd result: two strings with different thicknesses do not interact very much with each other. It is as though the strings were separated spatially. One can reinterpret the thickness of the string to be a new spatial coordinate that goes away from the boundary.

Thus, a thin boundary string is like a string close to the boundary, whereas a thick boundary string is like one far away from the boundary [*see box on preceding page*]. The extra coordinate is precisely the coordinate needed to describe motion within the four-dimensional anti–de Sitter spacetime! From the perspective of an observer in the spacetime, boundary strings of different thicknesses appear to be strings (all of them thin) at different radial locations. The number of colors on the boundary determines the size of the interior (the radius of the Escher-like sphere). To have a spacetime as large as the visible universe, the theory must have about 10^{60} colors.

It turns out that one type of gluon chain behaves in the four-dimensional spacetime as the graviton, the fundamental quantum particle of gravity. In this description, gravity in four dimensions is an emergent phenomenon arising from particle interactions in a gravityless, three-dimensional world. The presence of gravitons in the theory should come as no surprise. Thanks to work by John H. Schwarz of the California Institute of Technology and Jöel Scherk of the École Normale Supérieure and, independently, Tamiaki Yoneya of Hokkaido University in Japan, physicists have known since 1974 that string theories always give rise to quantum gravity. The strings formed by gluons are no exception, but the gravity operates in the higher-dimensional space.

Thus, the holographic correspondence is not just a wild new possibility for a quantum theory of gravity. Rather, in a fundamental way, it connects string theory, the most studied approach to quantum gravity, with theories of quarks and gluons, which are the cornerstone of particle physics. What is more, the holographic theory seems to provide some insight into the elusive exact equations of string theory. String theory was actually invented in the late 1960s for the purpose of describing strong interactions, but it was later abandoned (for that purpose) when the theory of chromodynamics entered the scene. The correspondence between string theory and chromodynamics implies that these early efforts were not misguided; the two descriptions are different faces of the same coin.

Varying the boundary chromodynamics theory by changing the details of how the boundary particles interact gives rise to an assortment of interior theories. The resulting interior theory can have only gravitational forces, or gravity plus some extra force such as the electromagnetic force, and so on. Unfortunately, we do not yet know of a boundary theory that gives rise to an interior theory that includes exactly the four forces we have in our universe.

I first conjectured that this holographic correspondence might hold for a specific theory (a simplified chromodynamics in a four-dimensional boundary spacetime) in 1997. This

ALFRED T. KAMAJIAN

immediately excited great interest from the string theory community. The conjecture was made more precise by Polyakov, Stephen S. Gubser and Igor R. Klebanov of Princeton, and Edward Witten of the Institute for Advanced Study in Princeton, N.J. Since then, many researchers have contributed to exploring the conjecture and generalizing it to other dimensions and other chromodynamics theories, providing mounting evidence that it is correct. So far, however, no example has been rigorously proved—the mathematics is too difficult.

So far no example of the holographic correspondence has been rigorously proved— **the mathematics is too difficult.**

Mysteries of Black Holes

HOW DOES THE holographic description of gravity help to explain aspects of black holes? Black holes are predicted to emit Hawking radiation, named after Stephen Hawking of the University of Cambridge, who discovered this result. This radiation comes out of the black hole at a specific temperature. For all ordinary physical systems, a theory called statistical mechanics explains temperature in terms of the motion of the microscopic constituents. This theory explains the temperature of a glass of water or the temperature of the sun. What about the temperature of a black hole? To understand it, we would need to know what the microscopic constituents of the black hole are and how they behave. Only a theory of quantum gravity can tell us that.

Some aspects of the thermodynamics of black holes have raised doubts as to whether a quantum-mechanical theory of gravity could be developed at all. It seemed as if quantum mechanics itself might break down in the face of effects taking place in black holes. For a black hole in an anti–de Sitter spacetime, we now know that quantum mechanics remains intact, thanks to the boundary theory. Such a black hole corresponds to a configuration of particles on the boundary. The number of particles is very large, and they are all zipping around, so that theorists can apply the usual rules of statistical mechanics to compute the temperature. The result is the same as the temperature that Hawking computed by very different means, indicating that the results can be trusted. Most important, the boundary theory obeys the ordinary rules of quantum mechanics; no inconsistency arises.

Physicists have also used the holographic correspondence in the opposite direction—employing known properties of black holes in the interior spacetime to deduce the behavior of quarks and gluons at very high temperatures on the boundary. Dam Son of the University of Washington and his collaborators studied a quantity called the shear viscosity, which is small for a fluid that flows very easily and large for a substance more like molasses. They found that black holes have an extremely low shear viscosity—smaller than any known fluid. Because of the holographic equivalence, strongly interacting quarks and gluons at high temperatures should also have very low viscosity.

A test of this prediction comes from the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory in Upton, N.Y., which has been colliding gold nuclei at very high energies. A preliminary analysis of these experiments indicates the collisions are creating a fluid with very low viscosity. Even though Son and his co-workers studied a simplified version of chromodynamics, they seem to have come up with a property that is shared by the real world. Does this mean that RHIC is creating small five-dimensional black holes? It is really too early to tell. (Even if so, there is nothing to fear from these tiny black holes—they evaporate almost as fast as they are formed, and they "live" in five dimensions, not in our own four-dimensional world.)

Many questions about the holographic theories remain to be answered. In particular, does anything similar hold for a universe like ours in place of the anti–de Sitter space? A crucial aspect of anti–de Sitter space is that it has a boundary where time is well defined. The boundary has existed and will exist forever. An expanding universe, like ours, that comes from a big bang does not have such a well-behaved boundary. Consequently, it is not clear how to define a holographic theory for our universe; there is no convenient place to put the hologram.

An important lesson that one can draw from the holographic conjecture, however, is that quantum gravity, which has perplexed some of the best minds on the planet for decades, can be very simple when viewed in terms of the right variables. Let's hope we will soon find a simple description for the big bang! SA

MORE TO EXPLORE

Anti–de Sitter Space and Holography. Edward Witten in *Advances in Theoretical and Mathematical Physics,* Vol. 2, pages 253–291; 1998. Available online at **arxiv.org/abs/hep-th/9802150**

Gauge Theory Correlators from Non-Critical String Theory. S. Gubser, I. R. Klebanov and A. M. Polyakov in *Applied Physics Letters B,* Vol. 428, pages 105–114; 1998. **arxiv.org/abs/hep-th/9802109**

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

The Elegant Universe. Brian Greene. Reissue edition. W. W. Norton, 2003.

A string theory Web site is at **superstringtheory.com**

black hole COMPUTERS

In keeping with the spirit of the age, researchers can think of the laws of physics as computer programs and the universe as a computer

By Seth Lloyd and Y. Jack Ng

BLACK HOLE COMPUTER may sound absurd but is proving to be a useful conceptual tool for researchers studying cosmology and fundamental physics. And if physicists are able to create black holes in particle accelerators—as some predict will be possible within a decade—they may actually observe them perform computation.

JEAN-FRANCOIS PODEVIN

What is the difference between a computer and a black hole? This question sounds like the start of a Microsoft joke, but it is one of the most profound problems in physics today. Most people think of computers as specialized gizmos: streamlined boxes sitting on a desk or fingernail-size chips embedded in high-tech coffeepots. But to a physicist, all physical systems are computers. Rocks, atom bombs and galaxies may not run Linux, but they, too, register and process information. Every electron, photon and other elementary particle stores bits of data, and every time two such particles interact, those bits are transformed. Physical existence and information content are inextricably linked. As physicist John A. Wheeler of Princeton University says, "It from bit."

Black holes might seem like the exception to the rule that everything computes. Inputting information into them presents no difficulty, but according to Einstein's general theory of relativity, getting information out is impossible. Matter that enters a hole is assimilated, the details of its composition lost irretrievably. In the 1970s Stephen Hawking of the University of Cambridge showed that when quantum mechanics is taken into account, black holes do have an output: they glow like a hot coal. In Hawking's analysis, this radiation is random, however. It carries no information about what went in. If an elephant fell in, an elephant's worth of energy would come out—but the energy would be a hodgepodge that could not be used, even in principle, to re-create the animal.

That apparent loss of information poses a serious conundrum, because the laws of quantum mechanics preserve information. So other scientists, including Leonard Susskind of Stanford University, John Preskill of the California Institute of Technology and Gerard 't Hooft of the University of Utrecht in the Netherlands, have argued that the outgoing radiation is not, in fact, random—that it is a processed form of the matter that falls in. In 2004 Hawking came around to their point of view. Black holes, too, compute.

Black holes are merely the most exotic example of the general principle that the universe registers and processes information. The principle itself is not new. In the 19th century the founders of statistical mechanics developed what would later be called information theory to explain the laws of thermodynamics. At first glance, thermodynamics and information theory are worlds apart: one was developed to describe steam engines, the other to optimize communications. Yet the thermodynamic quantity called entropy, which limits the ability of an engine to do useful work, turns out to be proportional to the number of bits registered by the positions and velocities of the molecules in a substance. The invention of quantum mechanics in the 20th century put this discovery on a firm quantitative foundation and introduced scientists to the remarkable concept of quantum information. The bits that make up the universe are quantum bits, or "qubits," with far richer properties than ordinary bits.

Analyzing the universe in terms of bits and bytes does not replace analyzing it in conventional terms such as force and energy, but it does uncover new and surprising facts. In the field of statistical mechanics, for example, it unknotted the paradox of Maxwell's demon, a contraption that seemed to allow for perpetual motion. In recent years, we and other physicists have been applying the same insights to cosmology and fundamental physics: the nature of black holes, the fine-scale structure of spacetime, the behavior of cosmic dark energy, the ultimate laws of nature. The universe is not just a giant computer; it is a giant quantum computer. As physicist Paola Zizzi of the University of Padua in Italy says, "It from qubit."

By preparing the material that falls into a black hole, a hacker could program it to perform any desired computation.

When Gigahertz Is Too Slow

THE CONFLUENCE of physics and information theory flows from the central maxim of quantum mechanics: at bottom, nature is discrete. A physical system can be described using a finite number of bits. Each particle in the system acts like the logic gate of a computer. Its spin "axis" can point in one of two directions, thereby encoding a bit, and can flip over, thereby performing a simple computational operation.

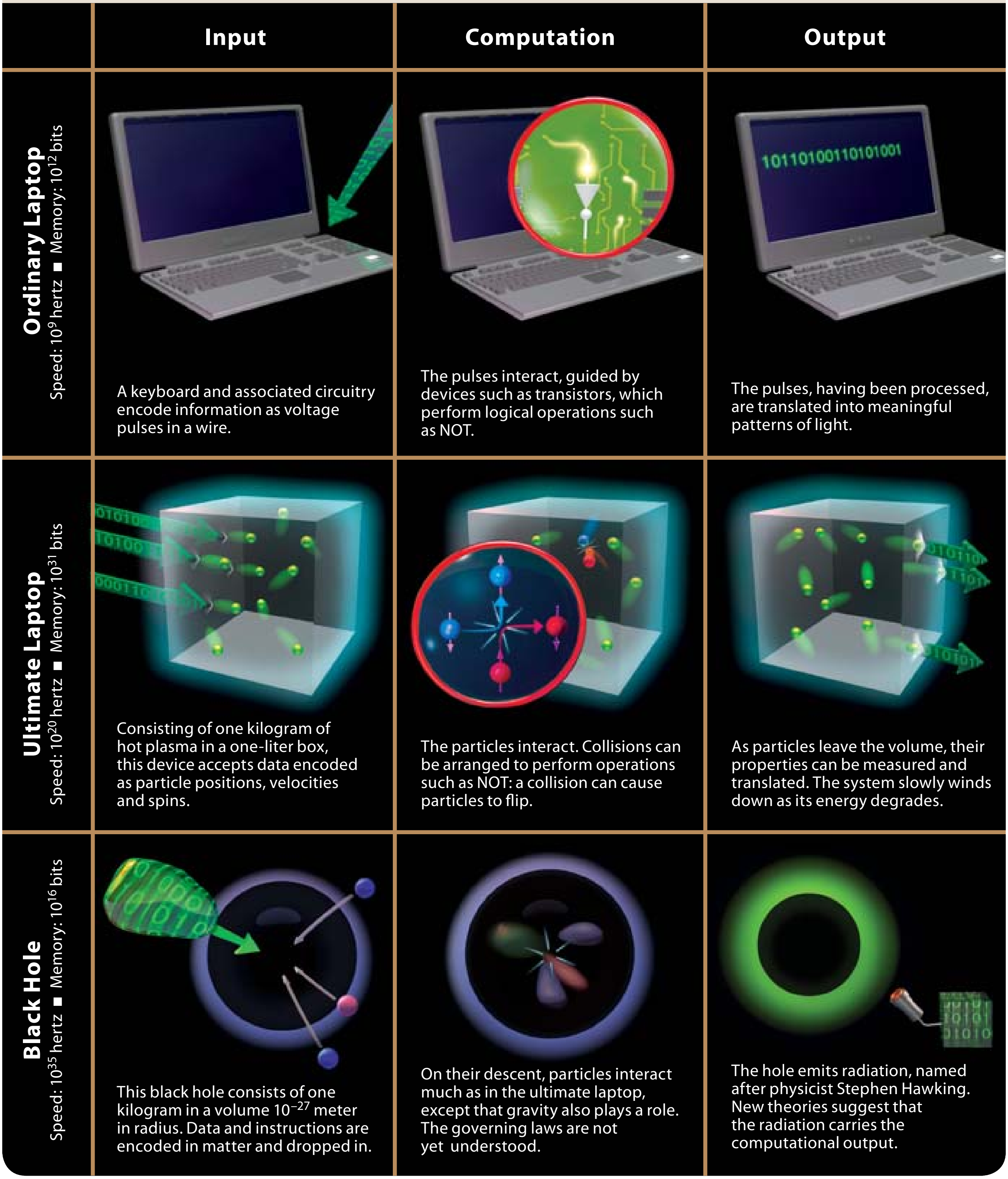
The system is also discrete in time. It takes a minimum amount of time to flip a bit. The exact amount is given by a theorem named after two pioneers of the physics of information processing, Norman Margolus of the Massachusetts Institute of Technology and Lev Levitin of Boston University. This theorem is related to the Heisenberg uncertainty principle, which describes the inherent trade-offs in measuring

Overview
Cosmic Computers

- Merely by existing, all physical systems store information. By evolving dynamically in time, they process that information. The universe computes.
- If information can escape from black holes, as most physicists now suspect, a black hole, too, computes. The size of its memory space is proportional to the square of its computation rate. The quantum-mechanical nature of information is responsible for this computational ability; without quantum effects, a black hole would destroy, rather than process, information.
- The laws of physics that limit the power of computers also determine the precision with which the geometry of spacetime can be measured. The precision is lower than physicists once thought, indicating that discrete "atoms" of space and time may be larger than expected.

Extreme Computing

What is a computer? That is a surprisingly complex question, but whatever precise definition one adopts, it is satisfied not just by the objects people commonly call "computers" but also by everything else in the world. Physical objects can solve a broad class of logic and mathematics problems, although they may not accept input or give output in a form that is meaningful to humans. Natural computers are inherently digital: they store data in discrete quantum states, such as the spin of elementary particles. Their instruction set is quantum physics.

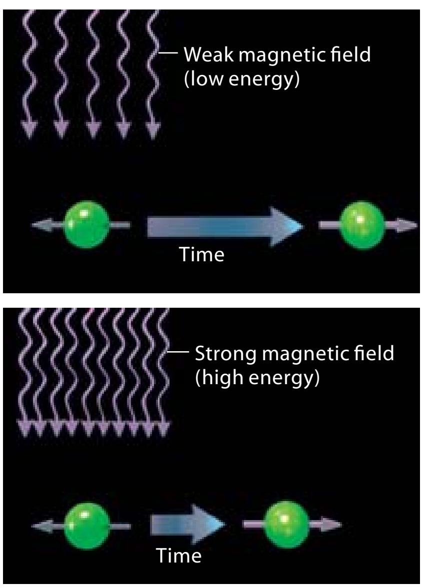


ALFRED T. KAMAJIAN

Understanding how information could leave a black hole is one of the liveliest questions in physics right now.

physical quantities, such as position and momentum or time and energy. The theorem says that the time it takes to flip a bit, t, depends on the amount of energy you apply, E. The more energy you apply, the shorter the time can be. Mathematically, the rule is $t \geq h/4E$, where h is Planck's constant, the main parameter of quantum theory. For example, one type of experimental quantum computer stores bits on protons and uses magnetic fields to flip them. The operations take place in the minimum time allowed by the Margolus-Levitin theorem.

From this theorem, a huge variety of conclusions can be drawn, from limits on the geometry of spacetime to the computational capacity of the universe as a whole. As a warm-up, consider the limits to the computational power of ordinary matter—in this case, one kilogram occupying the volume of one liter. We call this device the ultimate laptop.



FIRST LAW of quantum computation is that computation takes energy. The spin of a proton encodes a single bit, which can be inverted by applying a magnetic field. The stronger the field is—the more energy it applies—the faster the proton will flip.

ALFRED T. KAMAJIAN

Its battery is simply the matter itself, converted directly to energy per Einstein's famous formula $E = mc^2$. Putting all this energy into flipping bits, the computer can do 10^{51} operations per second, slowing down gradually as the energy degrades. The memory capacity of the machine can be calculated using thermodynamics. When one kilogram of matter is converted to energy in a liter volume, its temperature is one billion kelvins. Its entropy, which is proportional to the energy divided by the temperature, corresponds to 10^{31} bits of information. The ultimate laptop stores information in the microscopic motions and positions of the elementary particles zipping around inside it. Every single bit allowed by the laws of thermodynamics is put to use.

Whenever particles interact, they can cause one another to flip. This process can be thought of in terms of a programming language such as C or Java: the particles are the variables, and their interactions are operations such as addition. Each bit can flip 10^{20} times per second, equivalent to a clock speed of 100 giga-gigahertz. In fact, the system is too fast to be controlled by a central clock. The time it takes a bit to flip is approximately equal to the time it takes a signal to travel from one bit to its neighbor. Thus, the ultimate laptop is highly parallel: it acts not as a single processor but as a vast array of processors, each working almost independently and communicating its results to the others comparatively slowly.

In comparison, a conventional computer flips bits at about 10^9 times per second, stores about 10^{12} bits and contains a single processor. If Moore's law could be sustained, your descendants would be able to buy an ultimate laptop midway through the 23rd century. Engineers would have to find a way to exert precise control on the interactions of particles in a plasma hotter than the sun's core, and much of the communications bandwidth would be taken up in controlling the computer and dealing with errors. Engineers would also have to solve some knotty packaging problems.

In a sense, however, you can already purchase such a device, if you know the right people. A one-kilogram chunk of matter converted completely to energy—this is a working definition of a 20-megaton hydrogen bomb. An exploding nuclear weapon is processing a huge amount of information, its input given by its initial configuration and its output given by the radiation it emits.

From Nanotech to Xennotech

IF ANY CHUNK of matter is a computer, a black hole is nothing more or less than a computer compressed to its smallest possible size. As a computer shrinks, the gravitational force that its components exert on one another becomes stronger and eventually grows so intense that no material object can escape. The size of a black hole, called the Schwarzschild radius, is directly proportional to its mass.

A one-kilogram hole has a radius of about 10^{-27} meter, or one xennometer. (For comparison, a proton has a radius of 10^{-15} meter.) Shrinking the computer does not change its energy content, so it can perform 10^{51} operations per second, just as before. What does change is the memory capacity. When gravity is insignificant, the total storage capacity is proportional to the number of particles and thus to the volume. But when gravity dominates, it interconnects the particles, so collectively they are capable of storing less information. The total storage capacity of a black hole is proportional to its surface area. In the 1970s Hawking and Jacob D. Bekenstein of the Hebrew University of Jerusalem calculated that a one-kilogram black hole can register about 10^{16} bits—much less than the same computer before it was compressed.

In compensation, the black hole is a much faster processor. In fact, the amount of time it takes to flip a bit, 10^{-35} second, is equal to the amount of time it takes light to move from one side of the computer to the other. Thus, in contrast to the ultimate laptop, which is highly parallel, the black hole is a serial computer. It acts as a single unit.

How would a black hole computer work in practice? Input is not problematic: just encode the data in the form of matter or energy and throw them down the hole. By properly preparing the material that falls in, a hacker should be able to program the hole to perform any desired computation. Once the material enters a hole, it is gone for good; the so-called event horizon demarcates the point of no return. The plummeting particles interact with one another, performing computation for a finite time before reaching the center of the hole—the singularity—and ceasing to exist. What happens to matter as it gets squished together at the singularity depends on the details of quantum gravity, which are as yet unknown.

The output takes the form of Hawking radiation. A one-kilogram hole gives off Hawking radiation and, to conserve energy, decreases in mass, disappearing altogether in a mere 10^{-21} second. The peak wavelength of the radiation equals the radius of the hole; for a one-kilogram hole, it corresponds to extremely intense gamma rays. A particle detector can capture this radiation and decode it for human consumption.

Hawking's study of the radiation that bears his name is what overturned the conventional wisdom that black holes are objects from which nothing whatsoever can escape. The rate at which black holes radiate is inversely related to their size, so big black holes, such as those at the center of galaxies, lose energy much more slowly than they gobble up matter. In the future, however, experimenters may be able to create tiny holes in particle accelerators, and these holes should explode almost immediately in a burst of radiation. A black hole can be thought of not as a fixed object but as a transient congregation of matter that performs computation at the maximum rate possible.

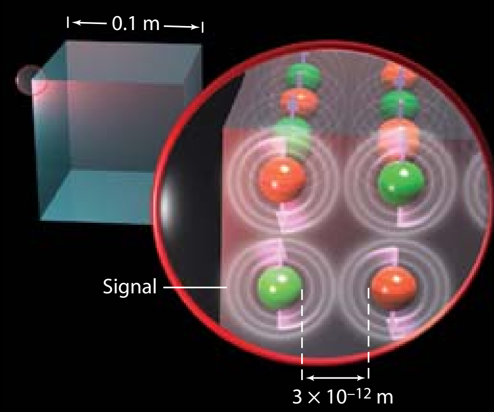
Escape Plan

THE REAL QUESTION is whether Hawking radiation returns the answer of the computation or merely gibberish. The issue remains contentious, but most physicists, including Hawking, now think that the radiation is a highly processed version of the information that went into the hole during its formation. Although matter cannot leave the hole, its information content can. Understanding precisely how is one of the liveliest questions in physics right now.

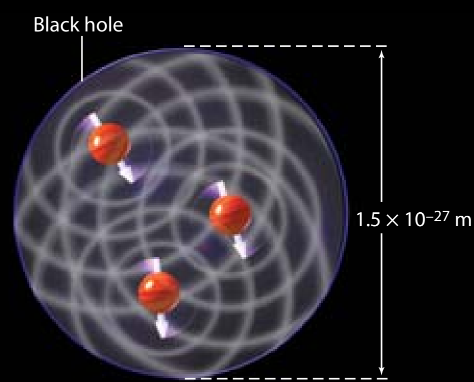
In 2003 Gary Horowitz of the University of California, Santa Barbara, and Juan Maldacena of the Institute for Advanced Study in Princeton, N.J., outlined one possible mechanism. The escape hatch is entanglement, a quantum phenomenon in which the properties of

Classifying Computers

The ultimate laptop and black hole computer embody two different approaches to increasing computing power. The ultimate laptop is the supreme parallel computer: an array of processors working simultaneously. The black hole is the supreme serial computer: a single processor executing instructions one at a time.



Ultimate laptop consists of a collection of particles that encode and process bits. Each can execute an instruction in 10^{-20} second. In that time, signals can move a distance of only 3×10^{-12} meter, which is roughly the spacing between particles. Therefore, communication is much slower than computation. Subregions of the computer work almost independently.

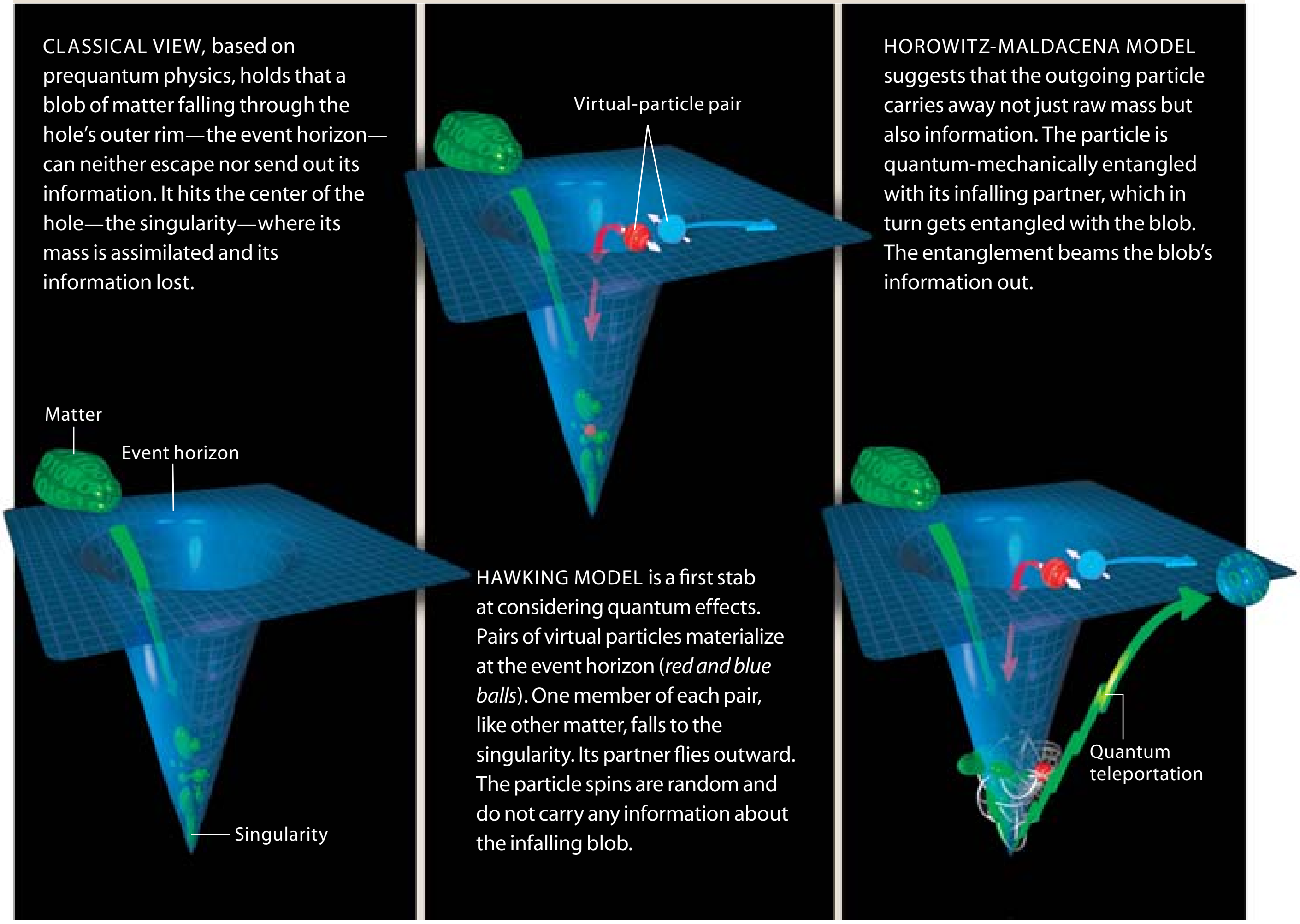


Black hole computer also consists of a collection of particles. Because of gravity, they encode fewer bits, giving more energy per bit. Each can execute an instruction in 10^{-35} second, which is the time it takes for a signal to cross the hole. Therefore, communication is as fast as computation. The computer operates as a single unit.

ALFRED T. KAMAJIAN

Evolution of Black Hole Theory

"Objects so dense that nothing, not even light, can escape"— this definition of black holes has become a cliché of newspaper articles and freshman astronomy lectures. But it is probably wrong. Physicists have argued since the mid-1970s that energy can leak out of a black hole, and most now think that information (which describes the form that the energy takes) can, too. These diagrams show a black hole from a hypothetical viewpoint outside spacetime.



CLASSICAL VIEW, based on prequantum physics, holds that a blob of matter falling through the hole's outer rim—the event horizon—can neither escape nor send out its information. It hits the center of the hole—the singularity—where its mass is assimilated and its information lost.

HAWKING MODEL is a first stab at considering quantum effects. Pairs of virtual particles materialize at the event horizon (*red and blue balls*). One member of each pair, like other matter, falls to the singularity. Its partner flies outward. The particle spins are random and do not carry any information about the infalling blob.

HOROWITZ-MALDACENA MODEL suggests that the outgoing particle carries away not just raw mass but also information. The particle is quantum-mechanically entangled with its infalling partner, which in turn gets entangled with the blob. The entanglement beams the blob's information out.

ALFRED T. KAMAJIAN

two or more systems remain correlated across the reaches of space and time. Entanglement enables teleportation, in which information is transferred from one particle to another with such fidelity that the particle has effectively been beamed from one location to another at up to the speed of light.

The teleportation procedure, which has been demonstrated in the laboratory, first requires that two particles be entangled. Then a measurement is performed on one of the particles jointly with some matter that contains information to be teleported. The measurement erases the information from its original location, but because of entanglement, that information resides in an encoded form on the second particle, no matter how distant it may be. The information can be decoded using the results of the measurement as the key.

A similar procedure might work for black holes. Pairs of entangled photons materialize at the event horizon. One of the photons flies outward to become the Hawking radiation that an observer sees. The other falls in and hits the singularity together with the matter that formed the hole in the first place. The annihilation of the infalling photon acts as a measurement, transferring the information contained in the matter to the outgoing Hawking radiation.

The difference from laboratory teleportation is that the results of this "measurement" are not needed to decode the information that was teleported. Horowitz and Maldacena argued that the annihilation does not have a variety of possible outcomes—only one. An observer on the outside can calculate this unique outcome using basic physics and thereby unlock the information. It is this conjecture that falls outside the usual formulation of quantum mechanics. Though controversial, it is plausible. Just as the initial singularity at the start of the universe may have had only one possible state, so it is possible that the final singularities inside black holes have a unique state. In June 2004 one of us (Lloyd) showed that the Horowitz-Maldacena mechanism is robust; it does not depend on what exactly the final state is, as long as there is one. It still seems to lead to a small loss of information, however.

Other researchers have proposed escape mechanisms that also rely on weird quantum phenomena. In 1996 Andrew Strominger and Cumrun Vafa of Harvard University suggested that black holes are composite bodies made up of multidimensional structures called branes, which arise in string theory. Information falling into the black hole is stored in waves in the branes and can eventually leak out. In 2004 Samir Mathur of Ohio State University and his collaborators modeled a black hole as a giant tangle of strings. This "fuzzyball" acts as a repository of the information carried by things that fall into the black hole. It emits radiation that reflects this information. Hawking has argued that quantum fluctuations prevent a well-defined event horizon from ever forming. The jury is still out on all these ideas.

The universe has performed the **maximum possible number of operations** allowed by the laws of physics.

Cyberspacetime

THE PROPERTIES of black holes are inextricably intertwined with those of spacetime. Thus, if holes can be thought of as computers, so can spacetime itself. Quantum mechanics predicts that spacetime, like other physical systems, is discrete. Distances and time intervals cannot be measured to infinite precision; on small scales, spacetime is bubbly and foamy. The maximum amount of information that can be put into a region of space depends on how small the bits are, and they cannot be smaller than the foamy cells.

Physicists have long assumed that the size of these cells is the Planck length (l_P) of 10^{-35} meter, which is the distance at which both quantum fluctuations and gravitational effects are important. If so, the foamy nature of spacetime will always be too minuscule to observe. But as one of us (Ng) and Hendrik van Dam of the University of North Carolina at Chapel Hill and Frigyes Károlyházy of Eötvös Loránd University in Hungary have shown, the cells are actually much larger and, indeed, have no fixed size: the larger a region of spacetime, the larger its constituent cells. At first, this assertion may seem paradoxical—as though the atoms in an elephant were bigger than those in a mouse. In fact, Lloyd has derived it from the same laws that limit the power of computers.

The process of mapping the geometry of spacetime is a kind of computation, in which distances are gauged by transmitting and processing information. One way to do this is to fill a region of space with a swarm of Global Positioning System satellites, each containing a clock and a radio transmitter [*see box on next page*]. To measure a distance, a satellite sends a signal and times how long it takes to arrive. The precision of the measurement depends on how fast the clocks tick. Ticking is a computational operation, so its maximum rate is given by the Margolus-Levitin theorem: the time between ticks is inversely proportional to the energy.

The energy, in turn, is also limited. If you give the satellites too much energy or pack them too closely together, they will form a black hole and will no longer be able to participate in mapping. (The hole will still emit Hawking radiation, but that radiation has a wavelength the size of the hole itself and so is not useful for mapping features on a finer scale.) The maximum total energy of the constellation of satellites is proportional to the radius of the region being mapped.

Thus, the energy increases more slowly than the volume of the region does. As the region gets bigger, the cartographer faces an unavoidable tradeoff: reduce the density of satellites (so they are spaced farther apart) or reduce the energy available to each satellite (so that their clocks tick more slowly). Either way, the measurement becomes less precise. Mathematically, in the time it takes to map a region of radius R, the total number of ticks by all the satellites is R^2/l_P^2. If each satellite ticks precisely once during the mapping process, the satellites are spaced out by an average distance of $R^{1/3}l_P^{2/3}$. Shorter distances can be measured in one subregion but only at the expense of reduced precision in some other subregion. The argument applies even if space is expanding.

This formula gives the precision to which distances can be determined; it is applicable when the measurement apparatus is just on the verge of becom-

THE AUTHORS

SETH LLOYD and *Y. JACK NG* bridge the two most exciting fields of theoretical physics: quantum information theory and the quantum theory of gravity. Lloyd, professor of quantum-mechanical engineering at the Massachusetts Institute of Technology, designed the first feasible quantum computer. He works with various teams to construct and operate quantum computers and communications systems. Ng, professor of physics at the University of North Carolina at Chapel Hill, studies the fundamental nature of spacetime. He has proposed various ways to look for the quantum structure of spacetime experimentally. Both researchers say their most skeptical audience is their family. When Lloyd told his daughters that everything is made of bits, one responded bluntly: "You're wrong, Daddy. Everything is made of atoms, except light." Ng has lost credibility on the subject because he is always having to turn to his sons for help with his computer.

Computing Spacetime

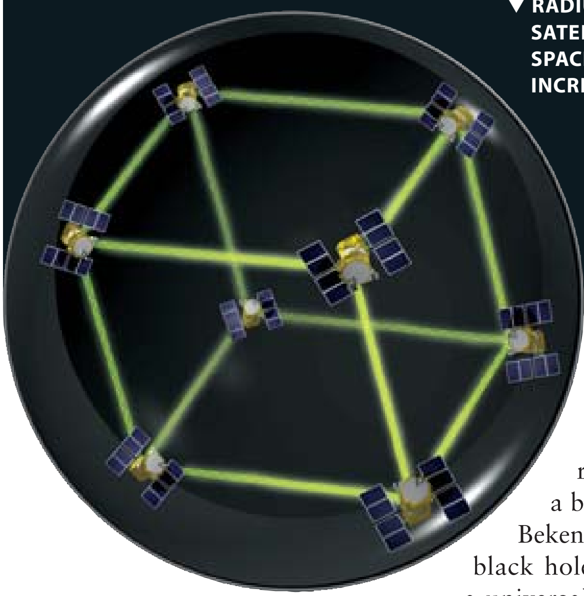
Measuring distances and time intervals is a type of computation and falls under the same constraints that computers do. It turns out that measurement is a much more slippery process than physicists had thought.

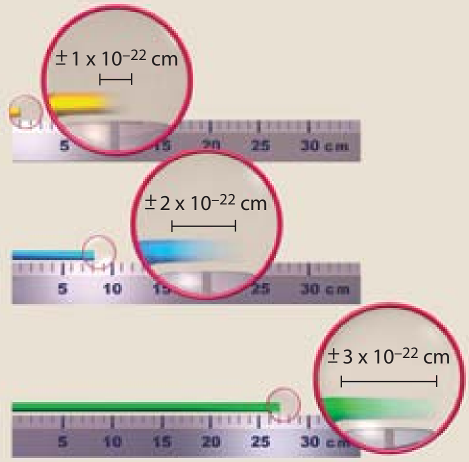
To map a volume of space, you might use a constellation of Global Positioning System satellites. They make measurements by sending signals and timing their arrival. For maximum precision, you need lots of satellites. But the number of satellites is limited: too many, and the entire system will collapse to a black hole.

To measure a region twice the size, you can use twice as many satellites. Because the volume is eight times as great, the satellites must be spaced farther apart. Each covers a larger subregion and can devote less attention to individual measurements, reducing their precision.

▲ **RADIUS: 100 km**
SATELLITES: 4
SPACING: 90 km

▼ **RADIUS: 200 km**
SATELLITES: 8
SPACING: 150 km
INCREASE IN ERROR: 26%





Measurement uncertainty is thus not fixed but can vary with the size of the object being measured. The larger the object is, the fuzzier its detailed structure. That differs from everyday life, in which the measurement imprecision is independent of the object and depends only on how finely subdivided your ruler is. It is as though your choice of what to measure affects the fine-scale structure of spacetime.

ALFRED T. KAMAJIAN

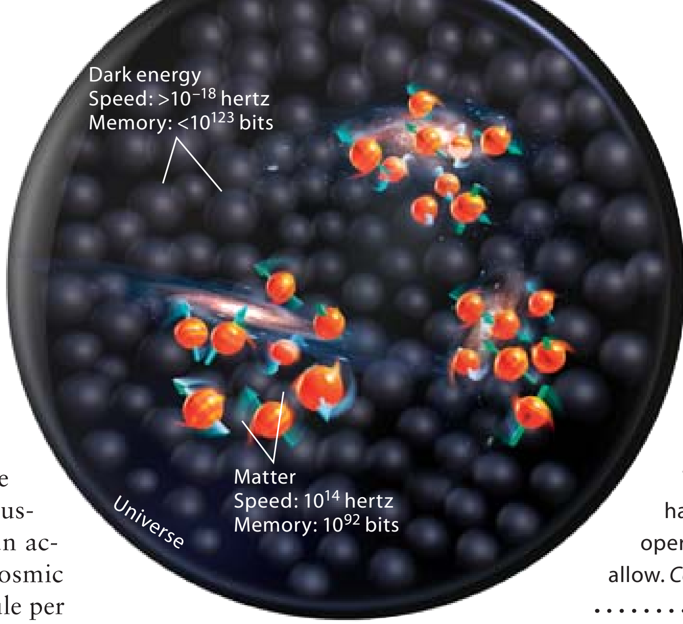
ing a black hole. Below the minimum scale, spacetime geometry ceases to exist. That level of precision is much, much bigger than the Planck length. To be sure, it is still very small. The average imprecision in measuring the size of the observable universe is about 10^{-15} meter. Nevertheless, such an imprecision might be detectable by precise distance-measuring equipment, such as future gravitational-wave observatories.

From a theorist's point of view, the broader significance of this result is that it provides a new way to look at black holes. Ng has shown that the strange scaling of spacetime fluctuations with the cube root of distances provides a back-door way to derive the Bekenstein-Hawking formula for black hole memory. It also implies a universal bound for all black hole computers: the number of bits in the memory is proportional to the square of the computation rate. The proportionality constant is Gh/c^5—mathematically demonstrating the linkage between information and the theories of special relativity (whose defining parameter is the speed of light, c), general relativity (the gravitational constant, G) and quantum mechanics (h).

Perhaps most significantly, the result leads directly to the holographic principle, which suggests that our three-dimensional universe is, in some deep but unfathomable way, two-dimensional. The maximum amount of information that any region of space can store seems to be proportional not to its volume but to its surface area. The holographic principle is normally thought to arise from the unknown details of quantum gravity, yet it also follows directly from the fundamental quantum limits to the precision of measurement.

The Answer Is . . . 42

THE PRINCIPLES of computation can be applied not just to the most compact computers (black holes) and the tiniest possible computers (spacetime foam) but also to the largest: the universe. The universe may well be infinite in extent, but it has existed a finite length of time, at least in its present form. The observable part is currently some tens of billions of light-years across. For us to know the results of a computation,



UNIVERSE IS A COMPUTER consisting of two types of components. Matter (*red*) is highly dynamic; it acts as a high-speed parallel computer. Dark energy (*gray*) appears to be nearly static; it acts as a lower-speed serial computer. Together the components have performed as many operations as the laws of physics allow. *Computo, ergo sum.*

ALFRED T. KAMAJIAN

it must have taken place within this expanse.

The above analysis of clock ticks also gives the number of operations that can have occurred in the universe since it began: 10^{123}. Compare this limit with the behavior of the matter around us—the visible matter, the dark matter and the so-called dark energy that is causing the universe to expand at an accelerated rate. The observed cosmic energy density is about 10^{-9} joule per cubic meter, so the universe contains 10^{72} joules of energy. According to the Margolus-Levitin theorem, it can perform up to 10^{106} operations per second, for a total of 10^{123} operations during its lifetime so far. In other words, the universe has performed the maximum possible number of operations allowed by the laws of physics.

To calculate the total memory capacity of conventional matter, such as atoms, one can apply the standard methods of statistical mechanics and cosmology. Matter can embody the most information when it is converted to energetic, massless particles, such as neutrinos or photons, whose entropy density is proportional to the cube of their temperature. The energy density of the particles (which determines the number of operations they can perform) goes as the fourth power of their temperature. Therefore, the total number of bits is just the number of operations raised to the three-fourths power. For the whole universe, that amounts to 10^{92} bits. If the particles contain some internal structure, the number of bits might be somewhat higher. These bits flip faster than they intercommunicate, so the conventional matter is a highly parallel computer, like the ultimate laptop and unlike the black hole.

As for dark energy, physicists do not know what it is, let alone how to calculate how much information it can store. But the holographic principle implies that the universe can store a maximum of 10^{123} bits—nearly the same as the total number of operations. This approximate equality is not a coincidence. Our universe is close to its critical density. If it had been slightly more dense, it might have undergone gravitational collapse, just like the matter falling into a black hole. So it meets (or nearly meets) the conditions for maxing out the number of computations. That maximum number is $R^2/l_P{}^2$, which is the same as the number of bits given by the holographic principle. At each epoch in its history, the maximum number of bits that the universe can contain is approximately equal to the number of operations it could have performed up to that moment.

Whereas ordinary matter undergoes a huge number of operations, dark energy behaves quite differently. If it encodes the maximum number of bits allowed by the holographic principle, then the overwhelming majority of those bits have had time to flip no more than once over the course of cosmic history. So these unconventional bits are mere spectators to the computations performed at much higher speeds by the smaller number of conventional bits. Whatever the dark energy is, it is not doing very much computation. It does not have to. Supplying the missing mass of the universe and accelerating its expansion are simple tasks, computationally speaking.

What is the universe computing? As far as we can tell, it is not producing a single answer to a single question, like the giant Deep Thought computer in the science-fiction classic *The Hitchhiker's Guide to the Galaxy*. Instead the universe is computing itself. Powered by Standard Model software, the universe computes quantum fields, chemicals, bacteria, human beings, stars and galaxies. As it computes, it maps out its own spacetime geometry to the ultimate precision allowed by the laws of physics. Computation is existence.

These results spanning ordinary computers, black holes, spacetime foam and cosmology are testimony to the unity of nature. They demonstrate the conceptual interconnections of fundamental physics. Although physicists do not yet possess a full theory of quantum gravity, whatever that theory is, they know it is intimately connected with quantum information. It from qubit. SA

MORE TO EXPLORE

Ultimate Physical Limits to Computation. Seth Lloyd in *Nature*, Vol. 406, pages 1047–1054; August 31, 2000. Available online at **arxiv.org/abs/quant-ph/9908043**

From Computation to Black Holes and Space-Time Foam. Y. Jack Ng in *Physical Review Letters*, Vol. 86, No. 14, pages 2946–2949; April 2, 2001. Erratum, Vol. 88, No. 13, article 139902(E); March 14, 2002. **gr-qc/0006105**

Computational Capacity of the Universe. Seth Lloyd in *Physical Review Letters*, Vol. 88, No. 23, article 237901Z; June 10, 2002. **quant-ph/0110141**

The Black Hole Final State. Gary T. Horowitz and Juan Maldacena in *Journal of High Energy Physics*, JHEP02(2004)008; 2004. **hep-th/0310281**

Information: The New Language of Science. Hans Christian von Baeyer. Harvard University Press, 2004.